Performance Analysis of Multimodal Biometric Fusion

PhD Thesis

Waheeda Almayyan

This thesis is submitted in partial fulfilment of the requirements for the Doctor of Philosophy, awarded by Software Technology Research Laboratory

Faculty of Technology

De Montfort University

England, United Kingdom
February, 2012

Faculty of Technology

De Montfort University

England, United Kingdom
February, 2012

Dedication

To my brother *Haythem*…

Without your sacrifice, support and encouragement there would never have been any chance for this thesis to happen…

With *all* the *love* ♥♥♥

Abstract

Biometrics is constantly evolving technology which has been widely used in many official and commercial identification applications. The increased concerns in security during recent years have essentially resulted in more attention being given to biometric-based authentication techniques. A biometric-based authentication is basically a pattern recognition problem which makes a personal identification decision in order to determine the authority based on specific physiological or behavioural features. Most biometric systems that are currently in use typically employ a single biometric trait. Such systems are called unibiometric systems. Despite considerable advances in recent years, there are still challenges in authentication based on a single biometric trait, such as noisy data, restricted degree of freedom, intra-class variability, non-universality, spoof attack and unacceptable error rates [15,80].

Some of the challenges can be handled by designing a multimodal biometric system. Multimodal biometric systems are those which utilise or are capable of utilising, more than one physiological or behavioural characteristic for enrolment, verification, or identification. A variety of multimodal biometrics strategies have been proposed and analysed in literature. In these works, the integration of various biometric features is suggested for achieving more accurate authentication rate. So far, most published work on multimodal biometric fusion techniques has dealt primarily with the fusion at the score matching level.

Here, we suggest a novel fusion approach of iris and online signature traits. Online signature and iris authentication techniques have been employed in a range of

biometric applications. Besides improving the accuracy, the fusion of biometrics has several advantages such as increasing population coverage, deterring spoofing activities and reducing enrolment failure. In this doctoral thesis, we make a first attempt to combine online signature and iris biometrics. We principally explore the fusion of iris and online signature biometrics and their potential application as biometric identifiers. To address this issue, investigations is carried out into the relative performance of several statistical data fusion techniques for integrating the information in both unimodal and multimodal biometrics. We compare the results of the multimodal approach with the results of the individual online signature and iris authentication approaches. This thesis describes research into the feature and decision fusion levels in multimodal biometrics.

This research is novel in the following five ways. First, the performance of the iris recognition is improved due to using dual-tree complex wavelet transform features and support vector machine. Second, the accuracy of the online signature recognition is greatly increased with less number of features by combining global features with Rough set. Third, a decision-level fusion scheme between iris and online signature is introduced using binary particle swarm optimization; its performance is better than the conventional feature-level scheme. Fourth, this research deploy the particle swarm optimization scheme as a feature selection technique to enhance the performance of online signature and iris accuracy rates by eliminating redundant and irrelevant information. Fifth, a hybrid-level fusion technique combined by using ensemble of classifiers and the AND rule offers significant improvements to the accuracy of the suggested multimodal biometrics system.

Declaration

I declare that the work described in my thesis is original work undertaken by me for the degree of Doctor of Philosophy, at Software Technology Research Laboratory (STRL), at De Montfort University, United Kingdom. No part of the material described in this thesis has been submitted for the award of any other degree or qualification in this or any other university or college of advanced education.

Acknowledgments

By now I've come to realize that the acknowledgements segment of a PhD thesis is the most important section of all.

First and foremost I would like to thank my supervisor Professor Hussain Zedan for his unlimited support, devoted guidance and keen encouragement throughout my graduate studies.

I am very grateful to my second supervisor Dr. Hala Own for guiding me all throughout this thesis, motivating me at times when I needed it the most and above all for being a thoughtful person.

I would like to thank my PhD colleagues in STRL, particularly my closest colleague Dr. Awatef Rahuma, for years of friendship and productive working environment.

I would like to take this opportunity to express my inexpressible gratitude to the State of Kuwait – The Public Authority of Applied Education and Training– this thesis was made possible by a generous grant from.

Finally, I wish to express my sincere gratitude towards my beloved mother, my dearest sisters and brothers for their undiluted love, affection and encouraging me to fulfil my dream.

Publications

2011

W. Al-Mayyan, H.S. Own, H. Zedan. Rough set approach to online signature identification. Digital Signal Processing, Vol.21(3), pp. 477-485 , 2011.

W. Almayyan, H.S. Own, R. Ramadan H. Zedan. A Multimodal Biometric Fusion Approach based on Binary Particle Optimization. Proceedings of AI-2011 Thirty-first SGAI International Conference on Artificial Intelligence, Cambridge, England,pp.139-152, 2011.

2010

H.S. Own, W. Al-Mayyan, H. Zedan. Biometric-Based Authentication System Using Rough Set Theory. Proceedings of the 7th international conference on Rough sets and current trends in computing, Springer-Verlag , pp.560-569, 2010.

W. Almayyan, H.S. Own, H. Zedan. Iris features extraction using dual-tree complex wavelet transform. International Conference of Soft Computing and Pattern Recognition (SoCPaR 2010), pp.18-22, 2010.

W. Almayyan, H.S. Own, H. Zedan. Information Fusion in Biometrics: A Case Study in Online Signature. Proceedings of The International Multi-Conference on Complexity, Informatics and Cybernetics: IMCIC, Orlando, USA ,2010.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BMP	Bayes Multilayer Perceptrons
BPSO	Binary Particle Swarm Optimization
CASIA	Chinese Academy of Sciences—Institute of Automation
CFS	Correlation-based Objective Selection
DCT	Discrete Cosine Transform
DT-CWT	Dual-Tree Complex Wavelet Transform
DTM	Dynamic Time Warping
ERR	Equal Error Rate
FAR	False Acceptance Rate
FRR	False Rejection Rate
FTC	Failure To Capture
FTE	Failure To Enrol
GA	Genetic Algorithms
GAR	Genuine Accept Rate
HMM	Hidden Markov Models
IONN	Input Oriented Neural Networks
k-NN	k-Nearest Neighbour
LDA	Linear Discriminant Analysis
LPD	Linear Programming Descriptor classifier
LVQ	Learning Neural Network
NIST	The National Institute of Standards and Technology
PCA	Principal Component Analysis
PIN	Personal Identification Number
PSO	Particle Swarm Optimization
RBFNN	Neural Networks with Radial Basis Function
ROC	Receiver Operating Characteristics curve
SFFS	Sequential Feed Forward Selection
SFS	Sequential Forward Selection
SIFT	Scale Invariant Feature Transform
SVM	Support Vector Machine
TDNN	Delay Neural Networks
WT	Wavelet Transform

Chapter 1

Introduction

Biometrics is constantly evolving technology which has been widely used in many official and commercial identification applications. A biometric-based authentication is basically a pattern recognition problem which makes a personal identification decision in order to determine the authority based on specific physiological or behavioural traits. Most biometric systems that are currently in operation typically utilise a single biometric trait. Such systems are called unibiometric systems. Regardless of the significant advances in biometrics over the last few years, there are still major challenges in obtaining consistent authentication decision through unimodal-biometric based authentication approaches.

1.1 Thesis Motivation

There are numerous reasons that motivate our interest in enhancing the performance of unimodal authentication approaches. First of all, biometric features are not exactly the same every time they are gathered. For instance, your voice is subject to change within the same day due to your emotional mode or health state. Moreover, no two fingerprint are ever exactly the same. The quality of fingerprint images maybe degraded as a result of physical problems such as dry, oily, dirty finger, dirty sensor surface, scars and other factors or simply because the user has positioned his/her finger on the fingerprint sensor in a different position.

Several limitations of unimodal biometric systems can be overcome by integrating multiple biometric traits, such as collecting voice and face or multiple fingers of the same person. Such systems, known as multimodal (or multibiomteric) recognition systems, are expected to be more reliable due to the existence of multiple and independent pieces of confirmation.

Multimodal approach relies mainly on fusing separate information from different modalities to provide complementary information to achieve more reliable recognition of individuals. For example, a common approach is to combine face and speech modalities to achieve a more trustworthy recognition decision. Four levels of fusion are possible when integrating data from two or more biometric sources. These levels are sensor, feature, matching score and decision levels.

Nevertheless, multibiometric systems have drawbacks when compared to unimodal biometric based systems. They are more expensive as they should require more computational and storage resources. In addition, they also require a large number of test samples and additional time for user enrolment which usually cause inconvenience to the user. Furthermore, the precision of any multibiometric system can be worsened if the integrating of various biometrics was not followed by a proper classification technique.

In our research, we principally limit ourselves to two modalities, namely, iris and online signature. To the best of our knowledge, there is no reported research work that combined iris and online signature. The main motive behind the selection of iris and online signature as the biometric features for building a multimodal biometric system is that signature is being used for person authentication in most of the daily applications since a long time and iris offers an excellent recognition performance when used as a biometric.

1.2 Thesis Scope and Research Questions

As the core of our work throughout this thesis revolves around examining whether the performance of a biometric-based authentication system can be improved through integrating complementary biometric features which comes primarily from two different and independent modalities. Therefore, the main aim of the research will be to investigate the effectiveness of the suggested fusion techniques for multimodal biometrics, with the following specific objectives:

- Explore existing multimodal approaches.
- Develop and evaluate online signature-based authentication approach.
- Develop and evaluate iris-based authentication approach.
- Develop multimodal authentication system based on the selected biometrics. This involves:
  - Study the effectiveness of fusion of online signature and iris biometrics into the various fusion approaches in both unimodal and multimodal biometrics thorough experimental investigation.
  - Compare between the effect of applying feature-level and decision-level fusion approaches.
  - Enhance the performance of the proposed multimodal system.

All in all, the purpose of this work is to investigate whether the performance of a biometric system can be improved by integrating complementary information which comes primarily from the selected modalities.

This research poses a fundamental question: *which fusion scheme can achieve the best performance and how much improvement can be gained from the applying the suggested fusion schemes?*

Toward this objective, our intention is to design and develop several fusion schemes at different fusion levels. Additionally, we will also tackle the complexity problem, in the sense that we will also raise the question whether it could be possible to reduce the dimension of the fusion feature space, through an appropriate selection procedure, while keeping the same level of performance.

1.3 Thesis Contributions

This thesis makes the following main original contributions.

A Novel Online Signature Authentication Approach

A novel online signature identification scheme based on global features and Rough set is proposed. The information is extracted as time functions of various dynamic properties of the signatures. A database of 2160 signatures from 108 subjects was built. Thirty-one features were identified and extracted from each signature. Different feature reduction approaches and classifiers were used to assess their suitability for this application. Rough set approach has resulted in a reduced set of nine features that were found to capture the essential characteristics required for signature identification. The reported results from several experiments demonstrate the suitability and effectiveness of the Rough set approach in the application of online signature identification. This research approach and results appears in our publications [5,7,135].

Iris Features Extraction using Dual-Tree Complex Wavelet Transform

Iris offers an excellent recognition performance when used as a biometric. Iris patterns are believed to be unique due to the complexity of the underlying the environmental and genetic processes that influence the generation of iris pattern. Segmenting iris area is a challenging task since the iris images can be occluded by eyelids or eyelashes. This will cause a significant difference between the intra- and inter-class comparisons. In this thesis we suggest a new iris segmentation technique based on minimizing the effect of the eyelids and eyelashes. The iris texture information was represented by applying the dual-tree complex wavelet transform to build the feature vector. The proposed innovative technique proofed to be computationally effective as well as feasible in term of recognition rates compared with other techniques. The combination between dual-tree complex wavelet transform and SVM classifier is promising. This research approach and results appears in our publication [4].

Hybrid Fusion: Combining Feature and Decision-Levels

The objective of this work is to investigate the integration of online signature and iris features towards achieving a better performance that may not be achievable with single biometric. The experimental investigations have been concerned with the fusion of online signature and iris biometrics in the decision and hybrid fusion modes. The basic idea was to fuse and evaluate the decisions with the following set of well-known state-of-the-art-algorithms: SVM, Naïve Bayes and k-NN using fixed rules: Maximum, Sum, Majority and Minimum rules. The individual decisions from the two modalities were further combined with straightforward the AND logic rule to obtain the final decision. The AND logic was applied to ensure a satisfactory level of security, since a positive authentication is only accomplished in case if only all the fusion levels approaches produce positive authentication [80].

Taking advantage of both feature-level and decision-level fusions and in an attempt to improve the final authentication performance, we further proposed and developed a hybrid fusion technique. Based on the experimental results, it has been shown that the hybrid approach offers a considerable contribution to the accuracy of the suggested multimodal biometrics. This research approach and results appears in our publications [5,6].

1.4 Success Criteria

The success criteria of our research in this thesis are as follows:

- The research questions set at the beginning of this research have to be met,
- A study showing how the proposed architecture is improved upon existing tackled approaches,
- A study that shows the advantages of integrating online signature and iris features at a different fusion levels.

These criteria will be revisited in the conclusion chapter to argue that such solutions exist and met in our research.

1.5 Thesis Outline

The thesis is organised into seven chapters including this chapter. The chapter's organisation is illustrated in Figure 1.1. The content of each chapters are summarised as follows.

Chapter 2

We initially introduce the discipline of biometrics and its evolution towards multimodal biometrics. It investigates the key issues in multimodal biometric systems along with the different architectures for information integration, and a review of previous investigations in the literature related to multimodal biometrics.

Chapter 3

Present the results of authenticating online signature using global features. We describe our work on building an online signature database and performing statistical analysis of online signature signals. An online signature authentication algorithm based on comparing the performance of three feature selection algorithms was constructed for the selected feature set.

Chapter 4

Propose and develop a new segmentation approach for iris authentication based on minimizing the effect of the eyelids and eyelashes by trimming the iris area above the upper and the area below the lower boundaries of the pupil. The 2D dual-tree complex wavelet transform is extracted from the iris images and used to improve the recognition accuracy. The comparison of proposed features will be evaluated on a diverse classification schemes namely; Naïve Bayes, k-NN and SVM. The approach was evaluated on a benchmark iris dataset.

Chapter 5

This chapter investigate the possibility of fusing the information of iris image and online signature signals for the purpose of personal identification at the feature level. This chapter propose and investigate the usefulness of Binary Particle Swarm Optimization with a number of multimodal biometric scenarios.

Chapter 6

In this chapter an experimental investigation is conducted on the fusion of online signature and iris biometrics at the decision fusion mode. The basic idea was to fuse and evaluate the decisions of the SVM, Naïve Bayes and k-NN classifiers using fixed rules: Maximum, Sum, Majority and Minimum rules. The individual decisions from the two modalities were further combined with straightforward AND logic rule to obtain the final decision. Afterwards, taking advantage of both feature-level and decision-level fusions and in an attempt to improve the final authentication performance, we further proposed and developed a hybrid fusion technique.

Finally, **Chapter 7** contains the summary of our work and contributions made in this thesis and discuss directions for future work.

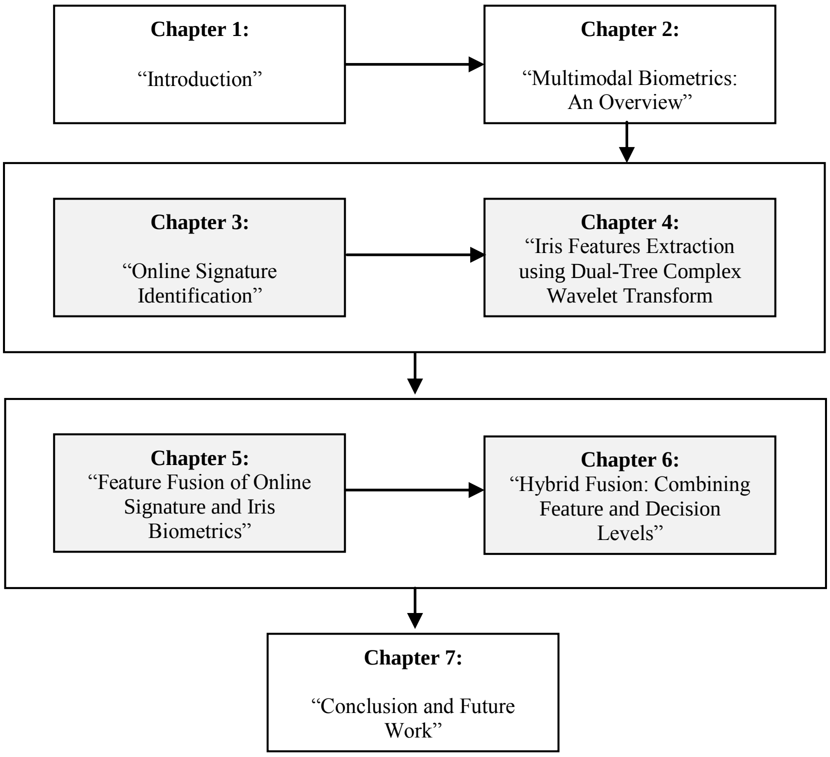


Figure 1.1 Outline of the thesis

Chapter 2

Multimodal Biometrics: An Overview

Derived from the ancient Greek words, "Bios" meaning life and "metron" meaning measures [93], biometric is defined as the statistical measurement and analysis of individual's physiological and/or behavioural distinctive features [80]. Biometrics has manifested itself as an efficient identity management system. Using parts of the human body as a mean to identity authentication goes back to ancient times. It is reported two thousand years ago that in ancient Babylon, merchants sealed deals with fingerprints on clay tablets to record their trading transactions [15]. The Chinese in the 3rd century B.C. used thumbprints and fingerprints on clay tablets as signatures to seal the official documents. While in the 14th century A.D., various official document papers dated in Persia bore fingerprint impressions [64,121].

A systematic and scientific basis for human identification started in the 19th century when a French police officer, Alphonse Bertillon [152] invented a number of anthropomorphic measurements, called Bertillonage, for identifying criminals. His system was built on the assumption that the body of people do not change in basic characteristics. Bertillon’s system involved measuring five primary measurements of body parts such as head length; head breadth; length of the middle finger and the length from elbow to end of middle finger (see Figure 2.1). Afterward, every major heading was additionally classified into three categories of: small, medium and large. The length of the little finger and the eye colour were also recorded.

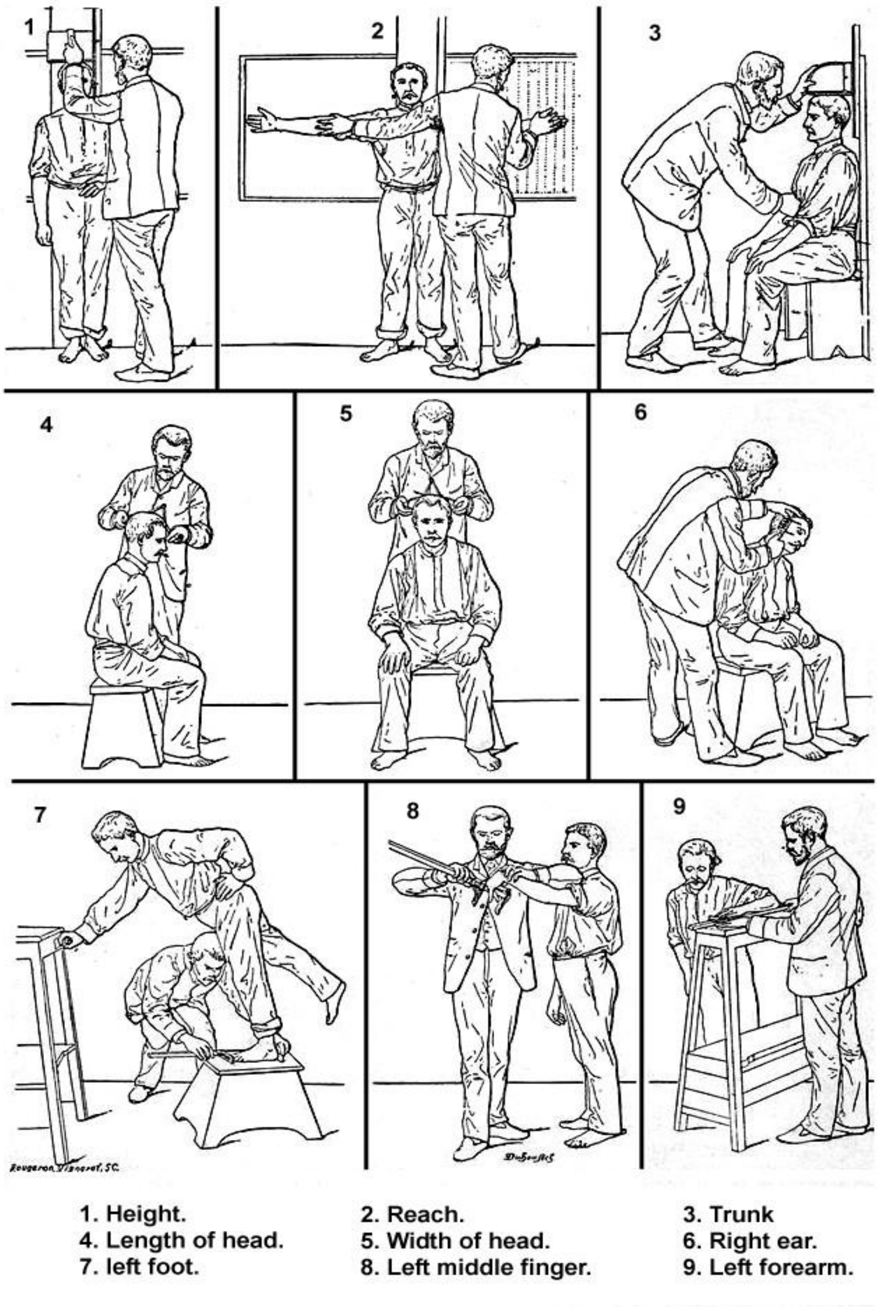


Figure 2.1 A chart from Bertillon's Identification anthropométrique (1893) demonstrating how to take measurements for his identification system, *adapted from* [152]

Biometrics-based personal authentication systems have recently gained intensive research interest due to the unreliability and inconvenience of traditional authentication systems. Biometrics recently became a vital component of any effective person identification solutions as biometric traits cannot be forged, shared, lost, duplicated, stolen or even forgotten [72,80,154].

Biometrics authenticates a user identity by the means of measuring an individual's unique physical or behavioural features. According to Maltoni et al. [116] these features can be classified into four categories static or dynamic and physical or knowledge-based biometrics as illustrated in Figure 2.2. Techniques that utilise the characteristic of fingerprints, palmprints or faces are considered static physical biometrics. Physical biometrics are related to the inherited physiological characteristics of the human body. Alternatively, behavioural biometric arise from activities carried out by that user either spontaneously or specifically learned. Dynamic or behavioural biometric techniques include, and not limited to handwritten signature, keystroke dynamics, gait patterns and lip movement. Techniques that use passwords or PINs (Personal Identification Number) are dynamic knowledge-based biometrics, whereas techniques that utilise magnetic cards and smart cards are considered static and physical-based biometrics.

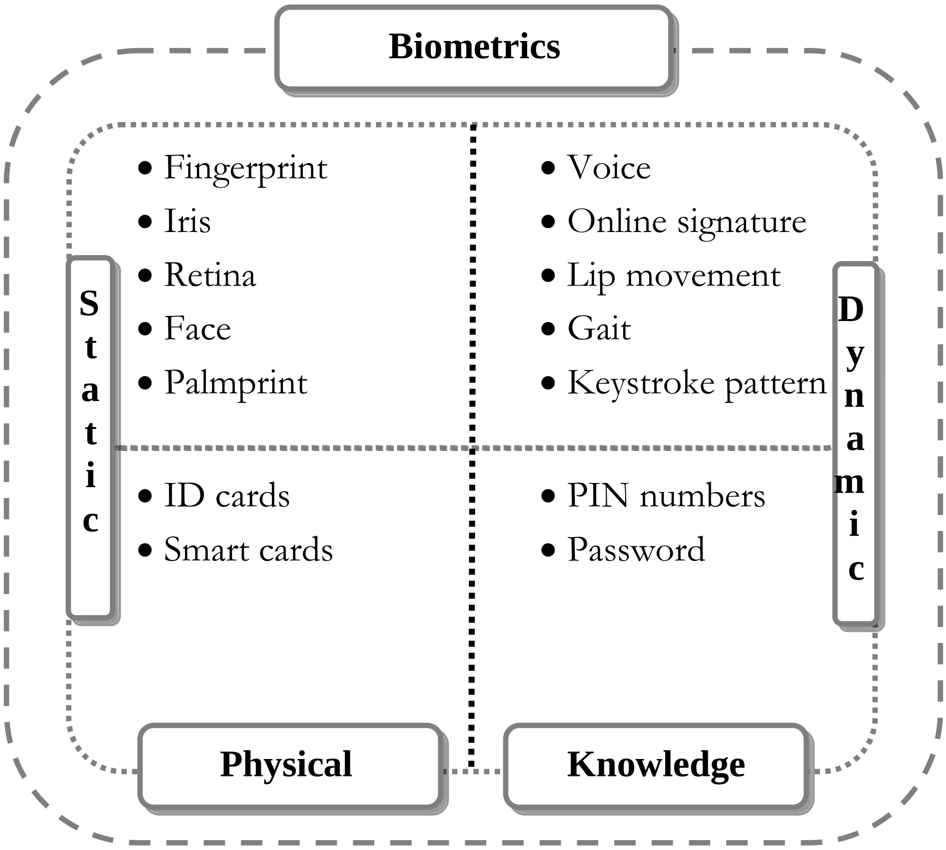


Figure 2.2 Authentication Techniques, *adapted from* [124]

Biometrics is a constantly growing technology which has been widely used in many successful official and commercial applications [72]. A biometric system is essentially a pattern recognition system which makes a personal identification decision by determining the authority of specific physiological or behavioural characteristics [80]. These are usually presented by the user when comparing biometric features with the stored feature of the claimed user.

2.1 Biometric Systems

Generally, any typical authentication biometric system comprises of the following units [72,80,116]:

- Data acquisition unit: consist of acquiring the biometric signal with a special sensor and converting it to a digital form.
- Feature extraction unit: extracts key information from the digital representation of the biometric cue.
- Matching unit: matches extracted features with templates stored in a database and output a similarity measure.
- Decision making unit: this final step issues a binary decision whether to accept or reject the claimed identity.

2.1.1 Biometric Recognition System Modes

Depending on the purpose behind its usage, biometrics can be used for identification or for verification. In verification mode, a user claims an identity and the system confirms his/her identity by comparing the biometric information submitted by the user with a reference for the claimed identity stored in the database. This is done by conducting a one-to-one comparison process.

The *verification* problem is in fact a two-category classification problem in the following manner [127]:
Given a feature vector X_Q and a claimed identity *I*, we need to determine if (I, X_Q) belongs to "*legitimate*" class denoted as ω_1 or "*impostor*" class denoted as ω_2. Let X_I be the stored template corresponding to identity *I*. In this case, X_Q is matched against X_I and a function that measures the similarity S and a pre-defined threshold η. Thus, the decision rule is given by

$$(I, X_Q) \in \begin{cases} \omega_1 & if \quad S(X_Q, X_I) \geq \eta \\ \\ \omega_2 & otherwise \end{cases} \qquad 2.1$$

Whereas in *identification* mode, the system compares the biometric information with all the templates stored in the database, in other words it is considered to be a one-to-many comparison. Given a feature vector X_Q and we need to determine the identity of I_k, $k=\{1,2,...,N\}$, where I_1, I_2,... I_N, are the classes enrolled in the system database. We need to determine if (I,X_Q) belongs to "*legitimate*" class or reject the sample if no correct class can be settled on. Thus,

$$X_Q \in \begin{cases} I_k & \text{if} \quad \max_k \{S(X_Q, X_{I_k})\} > \eta \\ \text{Re}\ ject & otherwise \end{cases} \qquad 2.2$$

Figure 2.3 illustrates the enrolment, identification and verification modules of a typical biometric system.

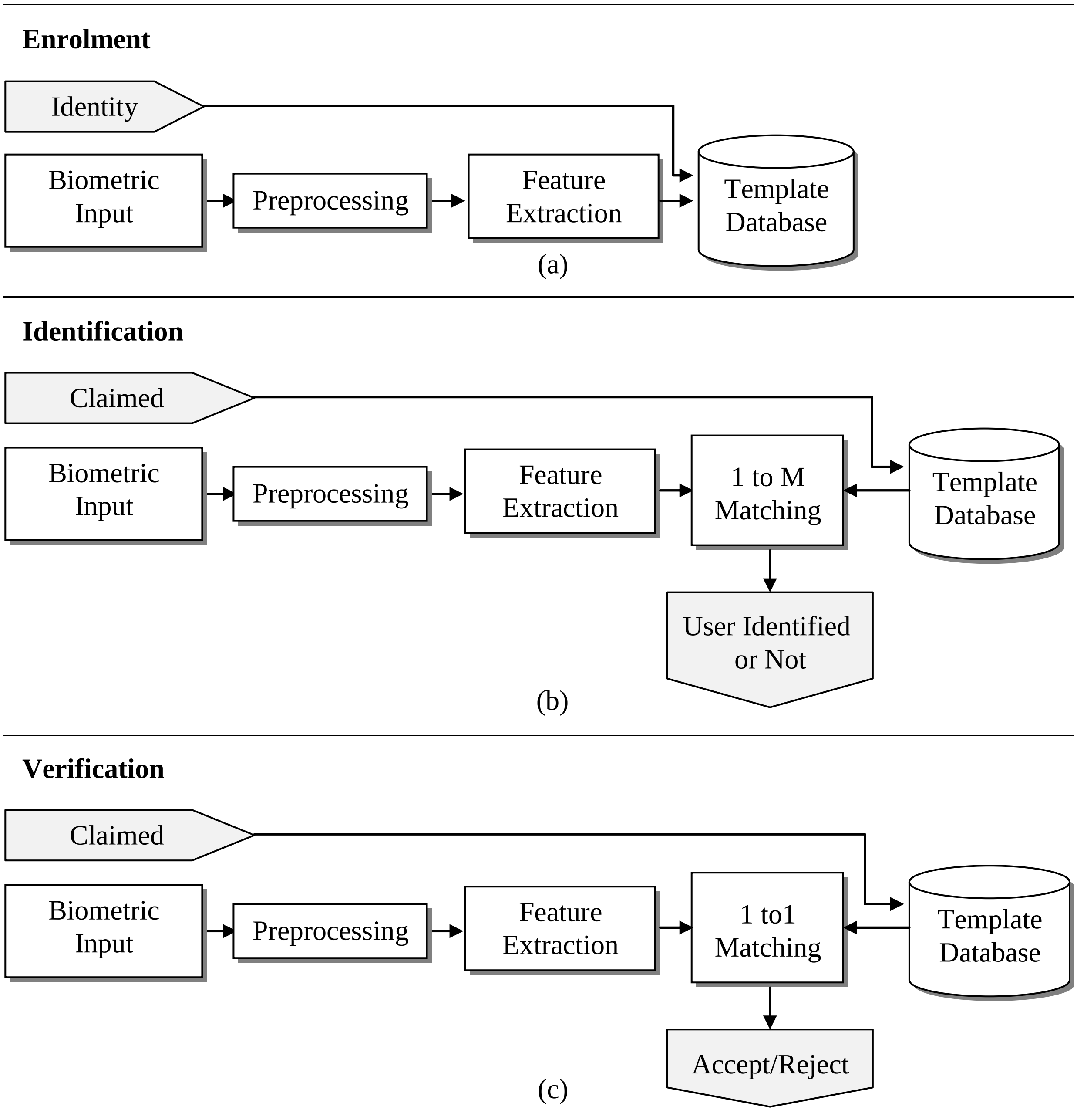


Figure 2.3 Diagram of the two modes of operation of a typical authentication system, (a) enrolment, (b) identification and (c) verification

2.1.2 Performance of a Biometric System

Unfortunately, biometric features are not exactly identical every time they are acquired. For example, your voice is subject to change within the same day due to your emotional mode or health state. Moreover, no two fingerprint are ever exactly the same. The quality of fingerprint images maybe degraded as a result of physical problems such as dry, oily, dirty finger, dirty sensor surface, scars and other factors or simply because the user has positioned his/her finger on the sensor in a slightly different position.

In evaluating the performance for any biometric based recognition system, there are mainly two types of factors: False Acceptance Rate (FAR) or type II and False Rejection Rate (FRR) also known as type I [72,154]. FAR is the probability that the system wrongly accept forged sample, while the FRR is the likelihood that a genuine access attempt will be unsuccessful. As these two factors are inversely related, lowering one of them often results in increasing the other, so it's common to describe the performance by another factor, the Equal Error Rate (ERR) where FAR equals FRR (Figure 2.4).

Another commonly used factor is Genuine Accept Rate (GAR) which is the probability that an authentic access will be accepted. Hence GAR = 1 – FRR, all these factors are dependent on the decision threshold T, and by varying decision threshold we can obtain multiple operating points of the system. The resulting plot of GAR against FAR is called the Receiver Operating Characteristics (ROC) curve, which is commonly used to evaluate the performance of biometric system.

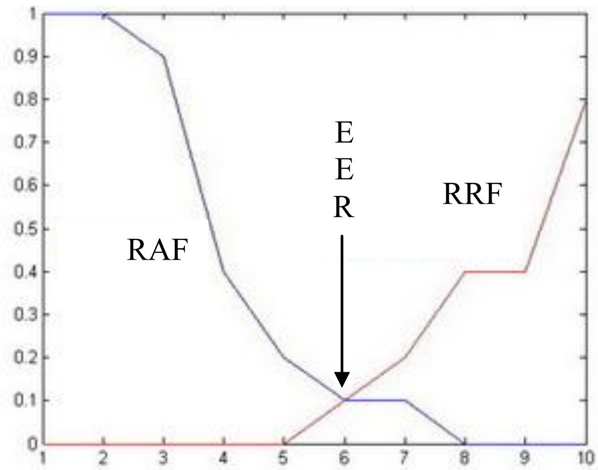


Figure 2.4 Typical FAR and FRR ROC curve

Other errors that may arise in a biometric system are Failure To Capture (FTC) and Failure To Enrol (FTE). These two errors are crucial for large-scale live applications. For instance, consider a scenario where passengers are authenticated using their fingerprint in an airport, in such a situation failure to capture the fingerprint data is problematic. The FTC error takes place when the data acquisition unit is not capable to capture a satisfactory quality of the biometric trait. As, if someone's voice is altered by cold he cannot be enrolled in a voiceprint recognition system. Whilst, the error of FTE usually occurs when the user tries to enrol in the recognition system are unsuccessful. Such as when the system rejects a fingerprint sample with poor image quality during enrolment as it do not match with the good quality templates stored in the database.

2.2 Multimodal Biometrics

Multimodal biometric systems are those which utilise, or are capable of utilising, more than one physiological or behavioural characteristic for enrolment either in verification or identification mode. It is generally believed that by integrating various biometric traits into one single unit, the limitations of unibiometric systems can be alleviated, given that several biometric sources usually compensate for the weaknesses of single biometric [68].

2.2.1 Limitations of unimodal biometric systems

In the last three decades, biometric based authentication systems such as fingerprint, palmprint, facial geometry, hand geometry, retinal and iris scans, signature recognition and voice recognition have being implemented in various applications including government IDs, computer and cellular phone logins, PDA, ATM, medical records management, border control, banking, e-commerce transactions and any place where identity management is critical [72, 80, 93,154,170,175]. Although, the successful implementation of biometric systems in these applications will still be constrained by the upper bound performance of the chosen biometric trait. Nevertheless, there is clearly a significant opportunity for improvement as suggested by the error rates shown in Table 2.1 which presents the state-of-the-art FRR and FAR of four popular biometrics. Clearly the accuracy rates rely on a number of test conditions such as the acquisition protocol, total number of subjects, number of biometric samples per subject, time lapse between data acquisition sessions, etc.

When looking for the potential biometric to be used in a specific identity authentication application, generally the following three criteria must be met [72,80,154]

Circumvent

- *Acceptability*, indicates people acceptance to use the biometric system.
- *Circumvention*, means how possible it is to fraud the authentication system.
- *Performance*, specifies the achievable identification (verification) accuracy and resources needed to achieve an acceptable accuracy.

Other factors may be less significant, such as

- *Universality*, means nearly all involved subjects should have or can produce the biometric.
- *Uniqueness*, means the difference between any two persons, should be sufficiently distinguishable.
- *Permanence*, which means the selected biometric should not change drastically under environment nor allow alteration.

In most cases, any physiological or behavioural characteristic that possess these properties can be used for personal identification. However, for the purpose of automatic personal identification, the biometric feature should have one additional property.

- *Collectability*, which means that the biometric should be quantitatively measurable.

Table 2.1 State-of-the-art false reject and false accept rates associated with fingerprint, face, voice, and iris verification systems, *adapted from* [154]

Biometric trait	Test	Test conditions	FRR	FAR
Fingerprint	FVC 2006	Heterogeneous population including manual workers and elderly people	2.2%	2.2%
	FpVTE 2003	US government operational data	0.1%	1%
Face	FRVT 2002	Controlled illumination, high resolution	0.8%–1.6%	0.1%
Voice	NIST 2004	Text independent, multi-lingual	5–10%	2–5%
Iris	ICE 2006	Controlled illumination, broad-quality range	1.1%–1.4%	0.1%

A brief comparison based on the above listed factors is provided in Table 2.2 based on the perception of the authors of [80]. Biometrics is considered to be a secure and convenient authentication means. Whereas some biometrics has gained more acceptance then others in range of applications, it is beyond doubt that utilising biometrics has gained a measure of acceptance. Nevertheless, each biometric modality has its strengths and limitations, and no single biometric modality is likely to meet all the desired performance of every authentication applications.

Table 2.2 Comparison of various biometric technologies based on the perception of the authors of [80].
Codes in the table H = high, M = medium, L = low

Biometric identifier	*Universality*	*Uniqueness*	*Permanence*	*Collectability*	*Performance*	*Acceptability*	*Circumvention*
DNA	H	H	H	L	H	L	L
Ear	M	M	H	M	M	H	M
Face	H	L	M	H	L	H	H
Facial thermogram	H	H	L	H	M	H	L
Fingerprint	M	H	H	M	H	M	M
Gait	M	L	L	H	L	H	M
Hand geometry	M	M	M	H	M	M	M
Hand vain	M	M	M	M	M	M	L
Voice	M	L	L	M	L	H	H
Keystroke	L	L	L	M	L	M	M
Odor	H	H	H	L	L	M	L
Palmprint	M	H	H	M	H	M	M
Retina	H	H	M	L	H	L	L
Signature	L	L	L	H	L	H	H
Iris	H	H	H	M	H	L	L

2.2.2 Motivation behind multimodal biometrics

The majority of currently in use biometric systems usually utilise a single biometric trait such systems are called unibiometric systems. Regardless of significant advances in the latest years, there are still several limitations derived from utilising one biometric trait. Such limitations should be considered before any real-time large-scale deployment projects. Some of the limitations are listed below [72, 80].

- **Noisy data acquired by sensor**

  This as a result of imperfect conditions or significant variation in the biometric itself during the biometric acquisition (see Figure 2.5). For example, a poorly illuminated face image may cause to incorrectly reject the subject's face sample. In fact, the recognition rate of any biometric system is very sensitive to biometric sample quality and noisy data can seriously reduce the overall accuracy of the system [25].

Figure 2.5 The negative effect of noisy images on fingerprint recognition.

The impression on the left is obtained from a subject during enrolment phase. The impression on the right is obtained from the same subject during verification phase after three months. Obviously, the development of scars or cuts can result in erroneous fingerprint matching results, *adapted from* [80]

- **Lack of universality**

A biometric modality is called universal as long as every subject of a target population is capable of presenting a valid biometric sample for authentication. This principle of universality is an essential condition in any efficient biometric recognition implementation. However, all biometric modalities are not really universal. The National Institute of Standards and Technology (NIST) has reported that it was not possible to acquire a good quality fingerprint from about 2% of the population (for instance, people with disabilities related to the hand, people with oily or dry fingertips, etc.) [132].
Consequently, such people cannot be signup in a fingerprint verification system. Therefore, errors occur during enrolment such as FTE and/or FTC is mostly related to using a single biometric feature.

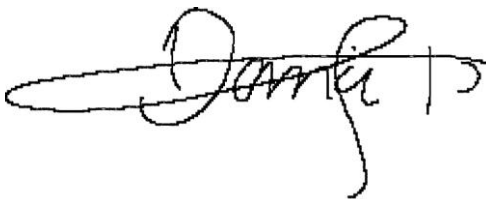
- **Lack of individuality or distinctiveness**

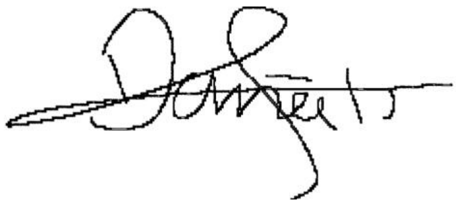
The biometric characteristics extracted from different persons may be quite similar. For instance, face recognition systems that depend on facial appearance fails in identifying identical twins. This short of distinctiveness usually increases the FAR of a biometric system.

- **Intra-class variations**

The biometric sample obtained from a user throughout the identification or verification phase is not identical to the sample which was collected to generate the reference database from the same user during the enrolment phase. This is known as the "intra-class" variations.
These variations may be to inappropriate interaction of the user with the sensor, as in the case when a user changes it pose or facial expression in front of a camera. This can happen when using different sensors at enrolment and

verification or due to alterations in the biometric modality, such as the case of developing new wrinkles in face or the presence of new scars in a fingerprint. Intra-class variations are more relevant in behavioural biometrics traits such as voice and signature (see Figure 2.6). Hence, individuals with large intra-class variability will regularly be falsely rejected as the acquired biometric trait they present does not match with any of the biometric template that they had enrolled with.


(a)


(b)


(c)

Figure 2.6 Signature intra-class variability.

(a), (b), and (c) are three signatures from a single user during one session

- **Sensitivity to attacks**

Many studies [3,82] demonstrated that it is possible to spoof a number of fingerprint authentication systems using simple techniques with molds made from range of materials such as plastic, clay, silicon or gelatine. Actually, behavioural biometric modalities are more susceptible to this kind of attack than physiological biometric modalities. Figure 2.7 shows some examples of real and faked fingerprints developed and tested in a recent research [52].

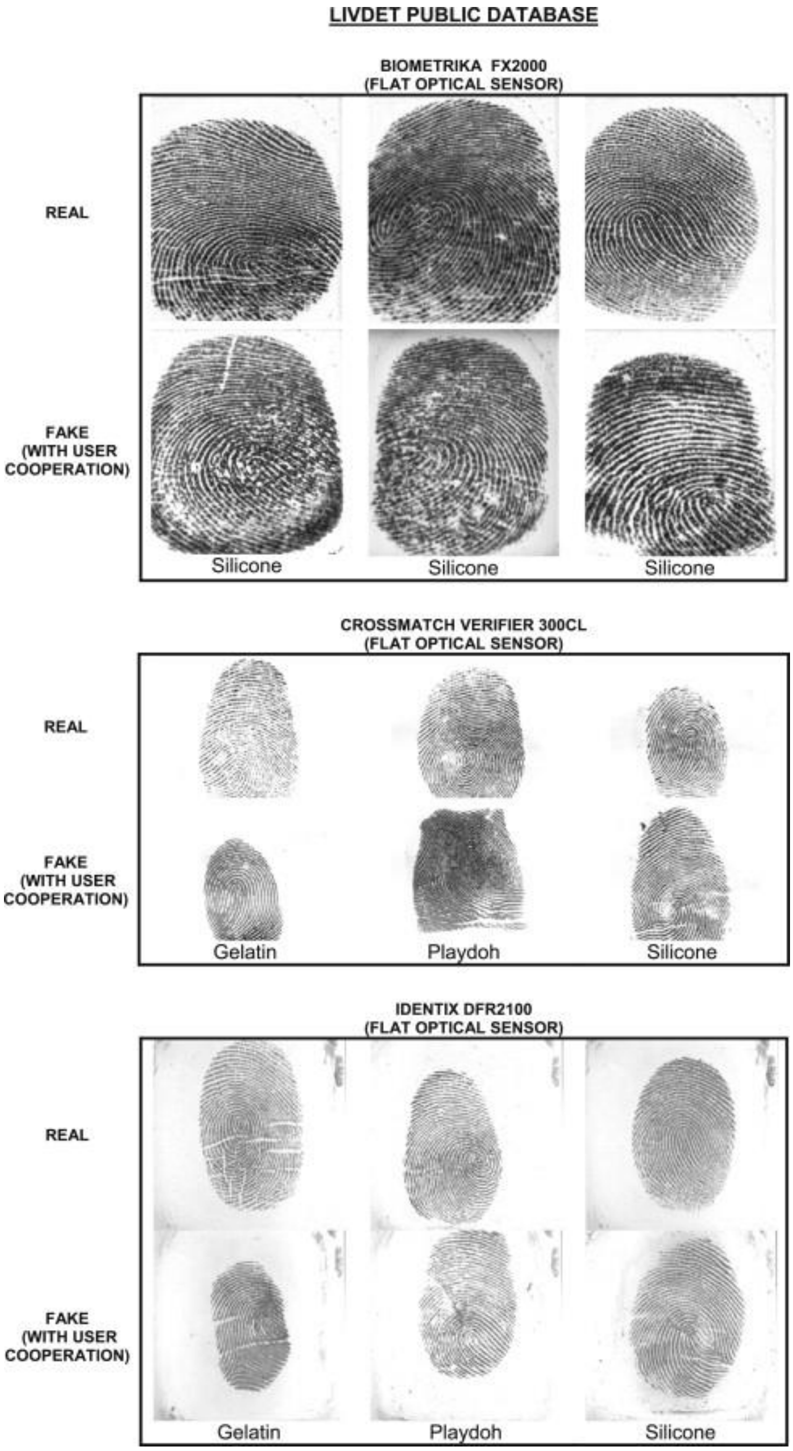


Figure 2.7 Typical examples of real and fake fingerprint images that can be found in the public database used in the experiments in [52]

Therefore, because of all these practical difficulties, the error rate associated with unimodal systems is relatively high. This makes unimodal-based authentication techniques improper for deployment in safety-critical or real-time applications. Some of aforementioned drawbacks can be overcome by considering a multimodal biometric approach. Multibiometric systems offer the following advantages over unibiometric systems:

1. Using an efficient fusion method to combine evidences from different sources can considerably improve the overall accuracy of the authentication system. Even though, the combination of several sources will enlarge the dimension size of the feature vector, it can decrease the overlap among the feature spaces of different classes [127].

2. Multibiometric systems are capable of addressing the problems related to unimodal biometrics such as non-universality. Thus, it can help in reducing the FRR and FAE and eventually improve the overall performance. For instance, if someone's voice is altered by cold he cannot be enrolled in a voiceprint recognition system, he can still be identified using other biometric traits like fingerprint or palmprint.

3. Multibiometric systems can add more flexibility to the enrolment procedure during user authentication. Lets us suppose a hypothetical access control application built using the modalities of face, voice and fingerprint. Later on, at the time of authentication, the user has the flexibility to choose all or a subset of available biometrics based on the nature of the application being considered and the convenience of the user. This is convenient for users with special needs, users with hand-related disabilities, for example, can enrol to the same system with their voice sample.

4. The noisy data, which usually have a considerable effect on the performance of the authentication process, can be considerably reduced with the availability of multiple sources of information. In such case, if the user failed to enrol using one of the sources due to acquisition conditions, he can try another biometric source.

5. Multibiometric systems have the capability to search a large scale template database in a computationally feasible way. This can be accomplished through using first the least accurate modality in pruning the database size down to a reasonable size before using the more accurate modality on the remaining database partitions.

6. Multimodal systems are more resistant to fraudulent techniques since it is not easy for an imposer to forge several biometric traits at the same time. By asking the subject to present the biometric traits in randomly order, the system can detect that the user is present at the acquisition point. To protect against spoofing and to ensure that only live traits are captured for enrolment or authentication, several studies suggested using "liveness detection" mechanism to measure the biometric trait physiological signs of life [52,82]. Liveness detection is an antispoofing technique ensures that only the biometric from a live subject is submitted for the purpose of authentication.

Nevertheless, multibiometric-based systems have drawbacks when compared to unibiometric-based systems. Unfortunately, they are more expensive as they should require more computational and storage resources. In addition, they also require a large number of test samples and additional time for user enrolment which usually cause inconvenience to the user. Furthermore, the precision of any multibiometric system can be worsened if the combining of the evidences was not followed by a proper classification technique.

2.3 Multimodal Biometric Fusion scenarios

A multibiometric system can be based on one or a combination of the biometric data obtainable from multiple sources. Any multibiometric system can be based on one or a combination of the following fusion scenarios [41,154]

- **Multiple modalities**

The biometric traits are extracted from two or multiple biometric modalities using single or multiple sensors. This is also known as multimodal biometrics. For example, a biometric recognition system based on combining face and ear attributes would be considered a multimodal system regardless of whether both images were captured by a different or the same imaging device.

- **Multiple sensors**

The same instance of biometric trait is obtained by different sensors. Such as in the case of verifying subject's face based on an image captured via two sources static digital image and video frame.

- **Multiple algorithms**

A single sample captured by a single sensor is processed by two or more different algorithms. For instance processing face recognition verification according to geometric (feature-based) or photometric (view-based) approaches is an example of processing multimodal biometrics using multiple algorithms.

- **Multiple instances**

A number of biometric samples from different instances of the same biometric trait is used in building such a system. An example of multiple instances is using left and right iris images for identity authentication. However, systems based on capturing sequential frames of face or ear images are considered to be multi-presentation rather than multi-instance.

- **Repeated instances**

The same biometric modality instance is acquired with the same sensor several times. As in capturing a sequential frame of facial images to construct a 3D facial image. This case is sometimes is not considered a multibiometric scenario.

Figure 2.8 Sources of multiple biometrics, *adapted from* [154]

To understand the distinction among the biometric fusion scenarios, table illustrates the basic distinctions among categories of multibiomteric implantation. The key aspect of the category that makes it "multi" is shown in boldface.

Table 2.3 Multibiomtrics category illustrated by the simplest case of using 2 of something, a*dapted from* [174]

Category	Modality	Algorithm	Biometric characteristic	Sensor
Multimodal	**2 (always)**	2 (always)	2 (always)	2 (usually)[a]
Multi-algorithmic	1 (always)	**2 (always)**	1 (always)	1 (always)
Multi-instance	1 (always)	1 (always)	**2 instance of 1 characteristic (always)**	1 (usually)[b]
Multi-sensorial	1 (always)	1 (usually)[c]	1 (always and same instance)	**2 (always)**
Multi-presentation	1	1	1	1

a - A multimodal system with a single sensor used to capture two different modalities (e.g. high resolution image used to extract face and iris or face and skin texture).

b - Exception may be the use of two individual sensors each capturing one instance (e.g. possibility a two fingerprint sensor).

c - It is possible that two samples from separate sensors could be processed by separate "feature extraction" algorithm, and then through a common comparison algorithm, making this "1.5 algorithms" or two completely different algorithms.

2.4 Multimodal Biometric Architecture

The next step after determining which biometric sources are to be integrated is to build the system architecture. Any multimodal system can operate in one of three different operational modes: serial, parallel or hierarchical mode [41,44].

- **Serial mode**

In this mode, sometimes called cascade mode, each modality is examined before the next modality is investigated. Therefore, multiple biometric traits do not have to be captured at the same time. Furthermore, a decision could be obtained before acquiring the rest of traits. As a result, the overall recognition duration can be decreased. For example in authentication system based on voice, fingerprint and iris traits. Initially the user uses the voice validation unit, and if this fails fingerprint validation is applied. If the last validation is failed the iris unit is required. The reward of such systems is that many users will enrol to the system using single trait.

- **Parallel mode**

In this mode of operation, the information from multiple modalities is processed concurrently, independently and all at once. Then the results are combined to make the final classification decision. Such as an authentication system based on voiceprint and face recognition. So, if it would be operated in a parallel mode, the user had to present the two traits in the same time for validation.

- **Hierarchical mode**

In this operational mode individual classifiers are combined in a treelike structure. This mode is preferred when a large number of classifiers are expected.

Most of the current multimodal biometric systems operate either in the serial mode or in the parallel mode. The serial mode is computationally efficient, whereas the parallel mode is more accurate [44].

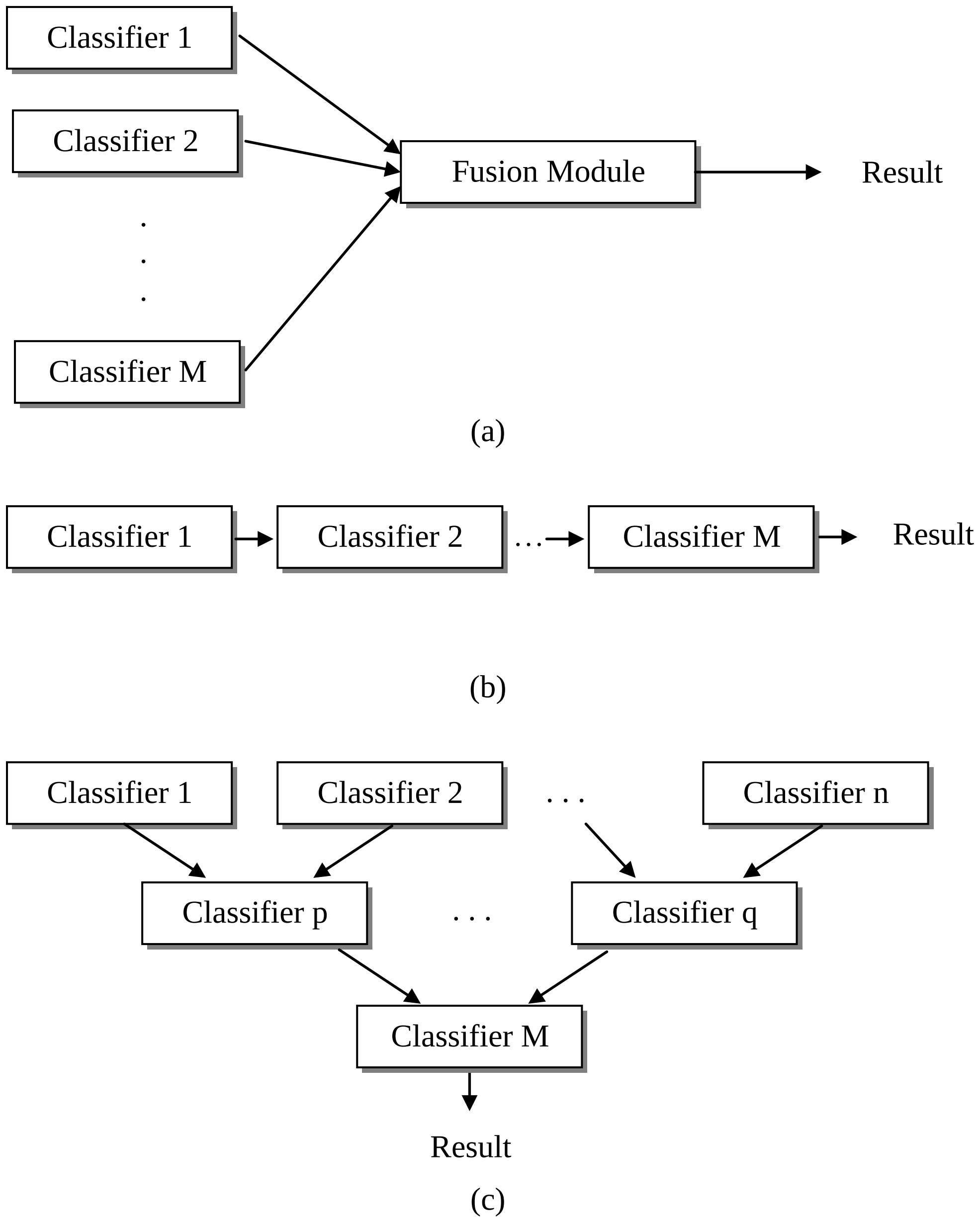


Figure 2.9 Architecture for several classifier combinations, *adapted from* [44], (a) parallel, (b) serial, (c) hierarchical

2.5 Multimodal Biometric Fusion levels

Most biometric-based authentication systems can be divided into four units: the sensor acquires the biometric data, the feature extraction unit process the biometric data in order to extract a discriminative representation of the acquired data. The matching unit, compares input features to stored templates, the decision unit issues either an “accept” or a “decline” decision based on the matching score.

Therefore, the fusion in multimodal systems can be performed at four potential levels: sensor, feature, matching and decision. The sensor and the feature levels are referred to as a *pre-mapping* fusion while the matching score and the decision levels are referred to as a *post-mapping* fusion [161]. Fusion levels are illustrated in Figure 2.10. In pre-mapping fusion, the biometric data are combined before classification. While in post-mapping fusion; each biometric data are modelled separately then all the biometric traits are combined after mapping into matching score/decision space. Pre-mapping schemes include fusion at the sensor and the feature levels. Whereas post-mapping schemes include fusion at the match score, rank and decision levels. The later approach has attracted a lot of attention although the amount of information available for fusion declined progressively after each layer of processing in a biometric system [41].

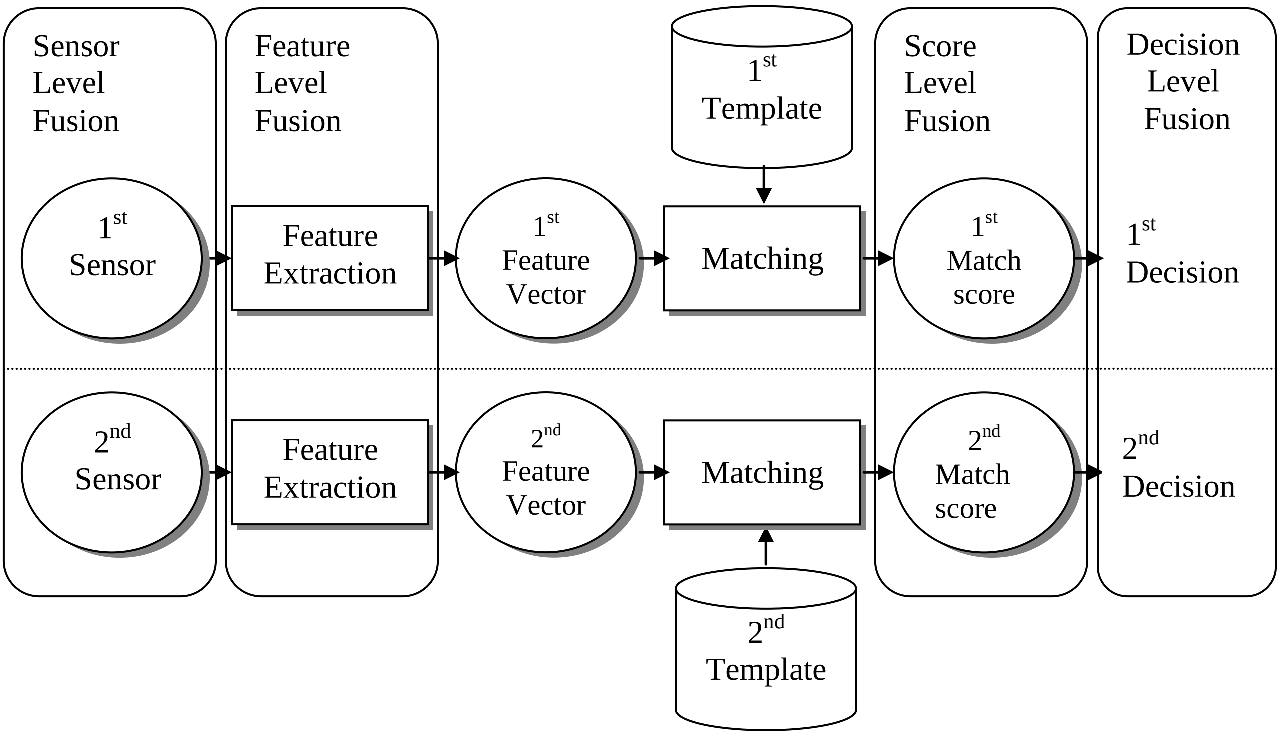


Figure 2.10 Fusion levels in multimodal biometric fusion, *adapted from* [128]

A. Pre-mapping fusion - Sensor level fusion

In this early stage of fusion, the raw data, derived from the same biometric characteristic with two or more sensors, is combined. Fusion at sensor level is closely associated with specific sensor types and a corresponding signal or image processing techniques. An example of the fusion at the sensor level is capturing a fingerprint image of each subject with two different sensors. Even though, fusion at primitive stages is expected to improve the recognition accuracy, it is not applicable with incompatible data gathered from different modalities.

Mosaicking has been investigated in face recognition methodologies. One approach proposed to model a statistical face model by constructing a mosaic from a video sequence of the face at various poses [112]. Another research [187] proposed an algorithm to construct a panoramic face using snapshots of five standard cameras that simultaneously acquire multiple views of a subject's face.

The fusion at the sensor level was a matter of research interests in fingerprint and face recognition. Constructing a composite fingerprint or face template using multiple impressions or 2D snapshots with the same sensor or camera, which is called Image mosaicking (or mosaicing), is a good example of sensor level fusion. Image mosaicking involves transforming and stitching of multiple images into a new collective image without any visible distortion in the overlapping areas [150]. Results indicated that mosaicking the fingerprint impressions first and then extracting the fingerprint minutiae templates, obtained a better matching performance [71,155,156]. A simple combination technique is applied in [22], where the normalized, masked ear and face images are concatenated to form a combined face ear image at the sensor level. The results show that fusion of more than one modality could lead to better results compared to the use of only one modality.

The resulting information from this initial level would potentially represent the richest source of information, whilst the other levels contain a smaller amount of information. Unfortunately, raw data may be corrupted by noise and may emphasize the intra-class variations.

B. Pre-mapping fusion - Feature level fusion

At this level, fusion can be applied to the extraction of different features from the same modality or different multimodalities to construct a joint feature vector, which then is utilised in matching and score modules. Merging extracted features into one single feature vector usually involves applying appropriate feature normalization, selection and reduction techniques [154]. Concatenating the feature vectors extracted from fingerprints and palmprint modalities are an example of a feature-level based system.

Since the feature level is certainly much richer and exploits more useful information about the raw biometric data, fusion at feature level is expected to perform better in contrast with fusion at score and decision levels [41]. Fusion at feature level may be helpful for closely-related modalities or for integrating features of the same modality with multiple sensors. However, such fusion type is not always feasible [154]. For example, in many approaches the given features might not be compatible due to differences in the nature of modalities. Moreover, the relationship between the feature spaces of the joint biometrics may not be known exactly. In addition, concatenating two feature sets or more may leads to the 'curse of dimensionality' problem. Furthermore, the majority of the practical commercial biometric systems do not provide access to the feature sets such as the raw fingerprint impressions of a fingerprint based commercial-of-the-self authentication systems.

C. Post-mapping fusion - Decision level fusion

In this approach, also denoted as abstract level, separate decisions taken from each biometric trait are combined at a very late stage. This seriously limits any efforts in enhancing the accuracy of the system through the fusion process. Thus, fusion at such a level is the least powerful [157]. Examples of combination techniques include AND rule, OR rule, majority voting, weighted voting based on Dempster-Shafer theory, etc.

D. Post-mapping fusion - Rank level fusion

This approach is possible only in identification systems where each classifier outputs a list of possible classes with rankings for each subject. The ranks of individual matchers are combined using techniques such as: the highest rank, Borda count and logistic regression approaches [103].

E. Post-mapping fusion - Matching score level fusion

At this level - also referred as decision, confidence, expert or opinion level- it is possible to combine scores obtained from the same biometric trait or different ones using one or more classifiers [158]. This fusion level can be divided into two categories: combination and classification. In the former approach, the separate matching scores are gathered to produce one score, which is used to make the final decision. In the latter approach, the input matching scores are considered as input features for a two-class pattern recognition problem, where the subject is classified as legitimate or not. The classifier presents a distance measure or a similarity measure between the input feature vector and the templates previously stored in the database. Before matching score fusion take place, normalization must be carried out.

There has been a proliferation of experimental studies trying to investigate the fusion of a range of biometric sources and examining different fusion techniques. The following table points to some of the representative work in the multibiometrics literature.

Table 2.4 Examples of multimodal systems

Modalities Fused	References	Level of Fusion
Face and voice	[90]	Match score
Face, voice and lip movement	[51]	Match score; Decision
Face, fingerprint and hand geometry	[158]	Match score
Face and iris	[127]	Feature
Face and gait	[81]	Match score
Face and ear	[24]	Sensor
Face and palmprint	[43]	Feature
Fingerprint, hand geometry and voice	[180]	Match score
Fingerprint and signature	[48]	Match score
Palmprint and hand geometry	[97]	Feature , Match score

2.6 Challenges related to Multimodal Biometric Systems Design

Multibiometric system design is certainly a challenging task since it is very difficult to choose the best possible sources of biometric information and fusion strategy for a particular application. This difficulty is related to many issues such as.

1. Benchmark multimodal datasets

The development of unimodal biometric databases of single-mode biometric traits has enabled the growth of unimodal biometric systems [184]. Yet, the development of multimodal systems is still limited because of the lack of consistent multi-biometric databases.

The number of the publicly accessible multimodal databases is quite limited. The main reason behind the creation of multimodal databases contents is that implies a certain degree of difficulty and challenges in the data acquisition phase. Furthermore, several controversial concerns are related to the legal and privacy of the data protection issue [184]. Moreover, most of the publicly available multimodal databases comprised of matching scores obtained by a number of biometric approaches operating on particular modalities [145, 45]. Consequently, this does not allow additional research to be held on other types of fusion levels other than the matching scores level. There are currently a few multimodal person authentication databases that are reported in the literature, some examples are listed in Table 2.5.

Table 2.5 List of available multimodal biometric databases, *adapted from* [37]

Database	Modalities
BANCA	Face and speech
XM2VTS	Face and speech
VidTIMIT	Face and speech
BIOMET	Biometric Score Set of: face, speech, fingerprint, hand and signature
NIST	Face and fingerprint
MYCT	Fingerprint and signature
UND	Face, ear profile and hand
FRGC	Face modality captured using camera at different angles and range sensors in different controlled or uncontrolled settings
IDIAP	Score of XM2VTS database
MyIDea	Face, speech, fingerprints, signature, handwriting, palmprint and hand Geometry
BioSec	Fingerprint, face, iris and voice

Due to the difficulties in constructing multimodal databases, some researchers have assumed that different biometric traits of the same person are statistically independent [158] in order to simplify the fusion algorithm design. Experiments in multimodal biometrics have been conducted on combining biometric trait of a user from a database with different biometric trait of another user from another different database to generate virtual or so-called chimeric databases [145].

2. Incompatibility of the information resources

As stated earlier, the integration of biometric information in early stages is thought to be more valuable since the amount of information available to the fusion module decreases as we move from one level of fusion to the next [72,116,127,154]. Nevertheless, fusion at early stages such as sensor and feature levels is not always possible due to the incompatibility of the gathered information. For example, in a multimodal biometric system based on fusing fingerprint and voiceprint, it is not possible to fuse the raw images of fingerprint with the voice signal.

3. Social acceptance and privacy issues

There is a number of serious privacy concerns raised concerning the implementation of biometrics, due to the fact that biometric technologies have the potential to provide governments and organizations with increased power over individuals. Privacy concerns are related to data collection, unauthorized use of recorded information and improper access to biometric records. As such, a trade-off between security concerns and privacy issues may be necessary by enforcing data protection laws and standards through common legislation [28]. Nevertheless, Biometrics from the positive point of view provides valuable tools to implement liable logs of system transactions.

4. Optimum design issues

The improvements in a multimodal biometric-based approach address key design questions [170]. The main question is about which modalities to integrate. This strongly depends upon the application and the required level of security concerns. This will also decides the complexity in designing the authentication system. Other design questions ought to be asked such as, What are the best combinations of modalities? How do we choose a best set of samples for a particular biometric? What is the smallest size sample set? Which level fusion is appropriate? Which is the best fusion scenario and processing architecture? What is expected performance? What is the cost involved in developing and deploying a real-time system?

Apart from the above mentioned factors there are still open questions to be addressed before deploying multimodal biometric system in a real time environment.

2.7 Summary

As the core of our work throughout this thesis revolves around fusing biometrics to improve the automatic authentication solution, we presented in this chapter a background about biometric and multimodal biometric. Biometrics recently became a significant part of any efficient person authentication solution as biometric traits cannot be stolen, shared or even forgotten. The majority of currently in use biometric systems usually utilise a single biometric feature, such systems are called unibiometric systems. Regardless of the significant advances in the field of biometric, there are still several limitations derived from utilising a single biometric trait.

Multimodal biometric systems are those which utilise, or are capable of utilising, more than one physiological or behavioural characteristic for enrolment either in the verification or identification mode. It is generally believed that by integrating various biometrics into one unit, the limitations of unibiometric systems can be alleviated, given that the several biometric sources usually compensate for the weaknesses of a single biometric. Four possible levels of fusion are used for integrating data from two or more biometric systems or sources. These levels are: sensor, feature, matching score and decision levels. Fusion at the feature level is an understudied problem.

In this thesis, we limit ourselves to iris and online signature modalities. To the best of our knowledge, there is no reported research work that combined iris and online signature. The main motive behind the selection of iris and online signature as biometric features for building a multimodal biometric system stems from their potential involvement in real-time large-scale biometrics applications.

The next two chapters discuss building online signature and iris unimodels in more details.

Chapter 3

Online Signature Identification

Biometrics-based authentication systems attracted a lot of attention as it is a promising alternative to password-based security systems. Handwritten signature as a distinctive personal biometric mark is considered among the more reliable biometrics. An essential advantage of the signature trait compared to other biometric characteristics is its longstanding traditional use as a method for identity verification. Compared to physical biometrics such as face, fingerprint and iris behavioural biometrics such as signature, voice and keystroke pattern change over time and thus have low intra-class variation. But usually physical biometrics requires special and quite expensive hardware to capture the biometric sample. While handwritten signature tend to vary slightly each time they are written and it is not quite as distinctive or hard to forge as finger or palm prints, the public's wide acceptance, nevertheless, makes it more suitable to be integrated into existing low-cost security authentication systems.

This chapter starts with introducing the problem of signature identification. Section 2 presents the literature review. The proposed system is presented and explained in Section 3 while the experimental results and analysis of results are described in Section 4. Finally, the overall summary will be given in the last section.

3.1 Signature as a Biometric

Handwritten signature authentication is the process of verifying the identity of a person based on his/her handwritten signature sample. Recognizing people by their handwritten signature has been an intense research area [45, 103,143] this is mainly due to the following factors [42].

- Signature is resistant to fraudulent access attempts. Even though, hypothetically, no person write his/her signature exactly the same each time, in practice, it is very difficult to forge the dynamic data (such as speed, pen-up movement, pressure, etc.) for every digitized signature point.

- Signature has been widely accepted as a means of legal and commercial transactions identity authentication. Signatures have played a historical role in authenticating documents. Being part of everyday life, signature based authentication is remarked as a consistent non-invasive authentication procedure by the majority of the users, therefore, it can help in overcoming some of the privacy difficulties [104,107].

- The user can be asked, if necessary, to change his/her signature. The main drawback of biometrics when compared with conventional methods is that many biometrics can be copied or forged [72,116,154]. Whereas it is always possible to obtain a new key or another password, it is not possible to replace any biometric data [73]. Nevertheless, signature is considered an exception where users can be asked to change their signature if needed.

Nevertheless, signature authentication is still a challenging issue for a number of reasons.

- Essentially, a signature reflects people's writing habits. Even though, some people may experience a lot of inconsistency between their signatures, mostly as a result of lack of signing habit. One possible solution to cope with this limitation is to acquire multiple signature instances during enrolment and not relying on a single instance. In addition, authentication should be conducted under similar conditions to those practiced during enrolment.

- While each ordinary literate human being has its unique style of writing his signature, yet signatures tend to evolve with time and the process of signing is influenced by physical and emotional state of the signatories.

3.1.1 Online signature recognition system

Signature identification and verification tasks fall into broad categories as either offline or online. In offline (*static*) systems, a signature is digitized with a scanner and only a static image record of the signature is stored. Thus, offline systems are of interest in situations where only hard copies of signatures are available. Whereas, online (*dynamic*) signature identification tracks down trajectory and other time-sequence variables such like velocity, pressure, etc. using specially designed tablets or other devices as the signature is being written. Given that online signatures also contain dynamic information, they are difficult to forge; it is appropriate for use in real-time applications, such as financial transactions, document authenticity and office automation.

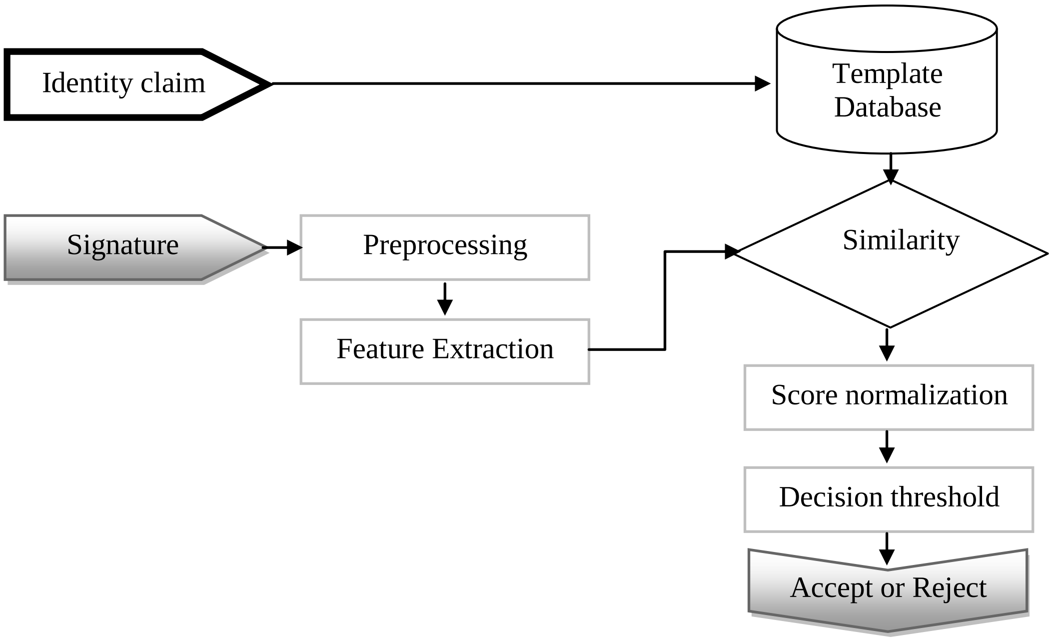


Figure 3.1 An architecture for an online signature identification system

The major modules of a typical online signature identification system involve.

- *Data acquisition* - recording the signatures trajectory and dynamics in addition to converting them to a digital form.
- *Signature preprocessing* - preprocess includes the acquired raw data either by translation, rotation or scaling if required and transforming the data into a standard format.
- *Feature extraction and building reference set* - extracting key information from the digital representation of the signature and create a signature reference set.
- *Classification* - matching extracted features with the reference templates stored in a database and output a fit ratio.

3.1.2 Online signature authentication techniques

There are two broad research methodologies for online signature-based biometric: function and parametric [37,103,143,144]. In the first methodology, the complete signals are considered as functions of time whose value directly constitutes the feature vector, such as position, pressure, acceleration, and velocity. While the parametric paradigm extracts local, global parameters or both from the entire signature trajectories and use it as statistical features or parameters, such as total signing time and number of zero crossing. In literature, several hundred parameters have been proposed for signature authentication. Overall, it has been established that functional methodology achieved better performance than parametric but they usually require more computational time [46,88,106,142].

Throughout the literature, different approaches and techniques have been developed for validating dynamic signature such as: feature values comparison, point to point comparison, Neural Network training techniques, Wavelets, Fourier Transform, Dynamic Time Warping (DTM), Hidden Markov Models (HMM), Vector Quantization, power spectral and shape comparison, etc.

The most widely studied online signature authentication technique is elastic matching using DTW. The purpose of DTW in signature authentication problems is to highlight interclass variability while suppressing intra-class variations. For online signature verification, DTW is a widely used technique to compare the similarity between online signature signals under test against templates disregarding the differences in time and speed. The winners of the 2004 first international Signature Verification Competition (SVC2004) [190] have used DTW to align signatures based on (Δx and Δy) local features [86]. Three reference sets were then calculated with based on the user's training set. Next, Principal Component Analysis (PCA) was performed to decorrelate the three distances and classify on this last measure. The algorithm was tested on the SVC2004 database, which consist of 20 authentic and 20 forged signatures gathered out of 40 persons. They achieved 1.65% FRR and 1.28% FAR with a database comprised of 94-users with a total of 182 authentic signatures and 313 skilled forgeries.
Nevertheless, there are still two main drawbacks of using DTW. It is computationally expensive and the resampling process usually results in losing important local details so that at the end forged signatures closely match genuine ones.

Another technique was motivated with the successful application of HMMs to speech recognition and online character recognition. HMMs have currently become the best performing statistical classifier for on-line signature verification [37]. In HMMs the similarity distance measure is actually the log-likelihood ratio of the acquired signature and the reference set.
Yang et al. [188] trained HMM to model the sequence of normalized angles along the trajectory of the signature. The model was tested on a database of 496 signatures gathered from 31 subjects. Their best result exhibited a FAR of 6.45% and a corresponding EER of 1.18%. Each signature was modelled by a single hiden-Markov model with left-to-right skip topology with 6-states. Each individual contributed 16 signatures, 8 were used for training and the rest 8 kept for testing. The results are given for random forgeries. For the individual HMM the Baum-

Welch algorithm was used for estimating the parameters of the HMM during training and testing. Shafiei and Rabiee [165] proposed a system based upon segmenting each signature based on its perceptually important points and then compute for each segment a feature vector comprised of seven features that are scale and displacement invariant, four of it are dynamic and three are static. The resulted vectors are used for training an HMM to achieve signature verification. With a database that included 622 genuine signatures and 1010 forgery signatures collected from a population of 69 subjects, the proposed system has achieved a FAR of 4% and a FRR of 12%.

Neural networks are known for their ability to solve complex functions by attempting to learn what the correct output should be from training data have been successfully applied in many pattern recognition problems, such as handwritten character recognition. Lee [105] has investigated the use of three neural network approaches for classifying signatures: Bayes Multilayer Perceptrons (BMP), Time Delay Neural Networks (TDNN), and Input Oriented Neural Networks (IONN). The input to the neural networks was a sequence of instantaneous absolute velocities extracted from the spatial coordinate. Consequently, the database used consists of 1000 genuine signatures from only one subject and 450 skilled forgeries from 18 trained forgers. The back propagation algorithm was used for network training. This experiment has shown that BMP provided the lowest misclassification error rate 2.67% among the three types of networks. Excellent results utilising neural networks were reported in [108]. They applied wavelets and back-propagation neural network together for the on-line signature verification purpose. They have used five feature functions to comprise the feature vector: the pen pressure, x and y velocity, angle of pen movement, and then applied the Daubechies-6 wavelet transform with 16 coefficients to compress the feature vector and using it at the end coefficients as input to a neural network. The system achieved a FRR of 0.0% and a FAR of less than 0.1% with a database of 922 genuine and forged signatures gathered from 41 subjects.
Zanuy [191] studied the performance of Vector Quantization, Nearest Neighbor, DTW, and HMM. A database of 330 users which includes 25 skilled forgeries performed by five different impostors has been used. Experimental results showed that the first proposed combination of VQ and DTW outperformed the other algorithms (DTW, HMM) and achieved a minimum detection cost function value equal to 1.37% for random forgeries and 5.42% for skilled forgeries.

Nanni and Lumini [129] proposed an on-line signature verification system based on local information and on a one-class classifier: the Linear Programming Descriptor classifier (LPD). The information was extracted as time functions of various dynamic properties of the signatures, then the discrete 1-D wavelet transform (WT) was performed on these features. The Discrete Cosine Transform (DCT) was used to reduce the approximation coefficients vector obtained by WT to a feature vector of a given dimension. Results using all the 5000 signatures from the 100 subjects of the SUBCORPUS-100 MCYT Bimodal Biometric Database [134] yielded an EER of 3–4% in the skilled forgeries and close to 1% in random forgeries.

In a recent research paper Nanni, Maiorana, Lumini and Campisi [130] developed a matching approach based on the fusion of Dynamic Time Warping, Hidden Markov Model and Linear Programming Descriptor classifiers. Furthermore, a template protection scheme employing the BioHashing and the BioConvolving approaches is discussed. The proposed system was tested with the same MCYT signature database and an EER of 3% was obtained when only five genuine signatures are acquired for each user during enrolment.

3.2 The proposed system

The proposed online signature approach consists of five main phases: Data acquisition, Pre-processing, Feature extraction and Feature reduction using Rough Set and Classification. Figure 3.2 depicts the overall framework of our proposed system. The process starts with acquiring the reference signature data with the help of a digitizing table to collect the dynamics of the signature. Then these signals are normalized to overcome the problem of different sizing in every signature. A feature vector is obtained to describe the global features of the signals. The Rough Set theory is applied to select the most significant features before classification.

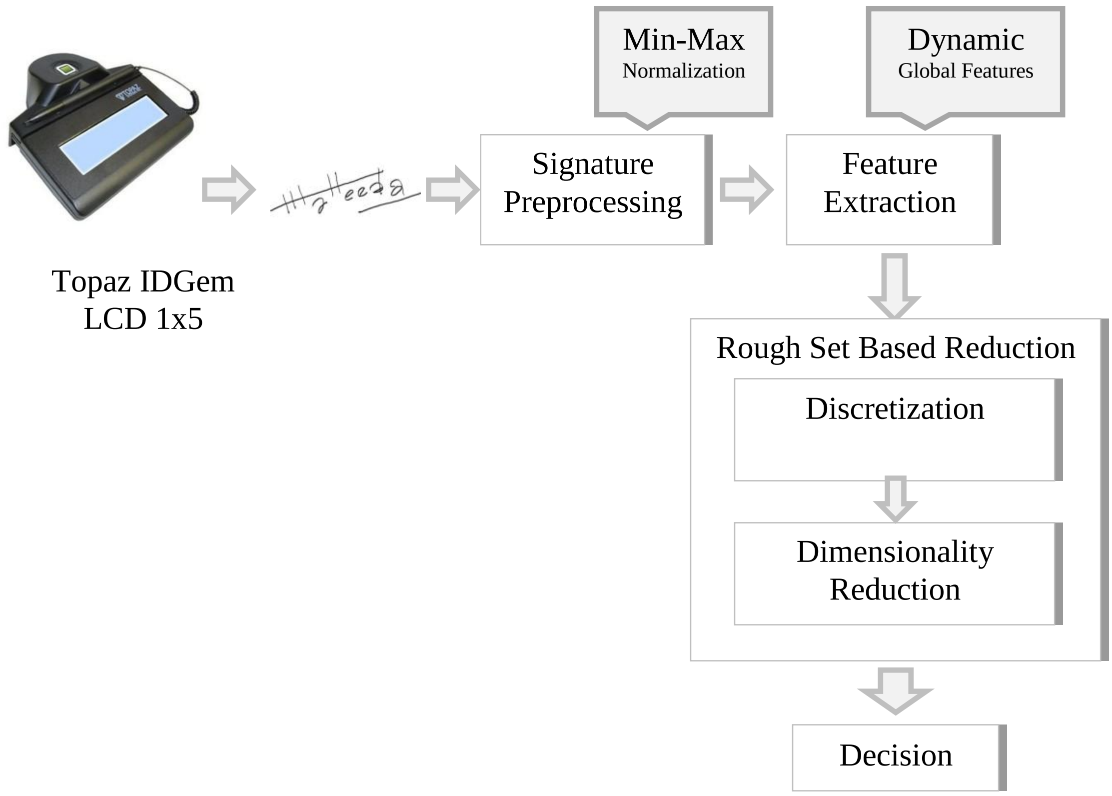
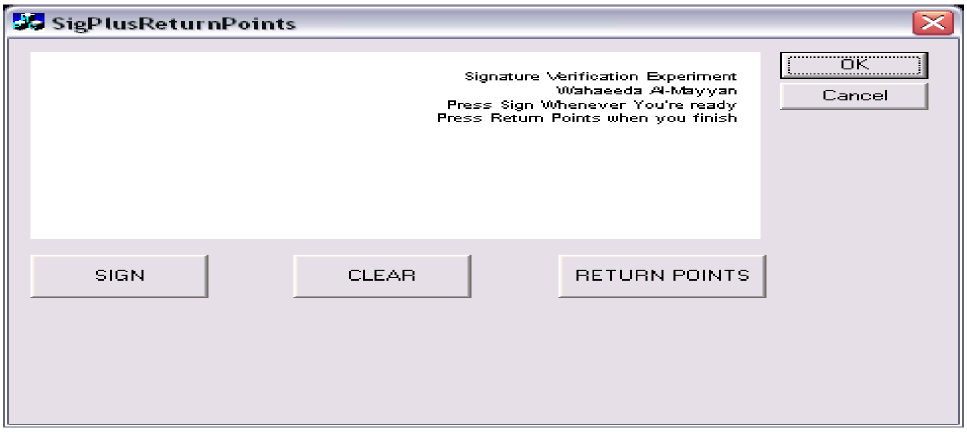


Figure 3.2 The overall framework of the proposed system

3.2.1 Data Acquisition

Online signatures can be captured using a variety of input devices such as digitizing tablets, specially designed pens, hand gloves [149] and tracking-camera. Overall, when one signs on a graphic tablet, two types of information are captured: the location coordinates and the timing information tagged to each pair of the x and y coordinates.

The proposed system implemented here uses Topaz's IdGem 1x5 signature pad, which is a non-sensitive pressure tablet with a visual feedback and LCD screen that gives the signer a natural feeling of signing on an ordinary paper [177]. The IdGem has a 4.4 x 1.3 inch effective writing area and captures samples at the rate of 377 points per second. The resolution is 410 true points per inch. The position values are translated into coordinates on the serial bus. The hardware interface is a Serial EIA Standard RS-232C port connected to a laptop computer running custom-written C++ driver software. Figure 3.3 illustrates the graphical user interface of the acquisition program.



(a)

(b)

Figure 3.3 The GUI of the developed signature capturing program,

(a) before (b) after signing

The values in the output stream produced by the digitiser are equidistant in time contain the following data:

- x(t), the x-coordinate sampled at timestamp t;
- y(t), the y-coordinate sampled at timestamp t.

In this approach, we restrict ourselves to features common to all digitizing tablet. At each sample point, we obtain the signature data as

$S(t) = [x(t)], y(t), timestamp(t)]^T$, $t = 1,\ldots,N$, where N is total the number of samples of the signature trajectory along with the timestamp and number of pen-ups are all recorded. Example of a signature and the function-based representation from the gathered signature database is depicted in Figure 3.4.

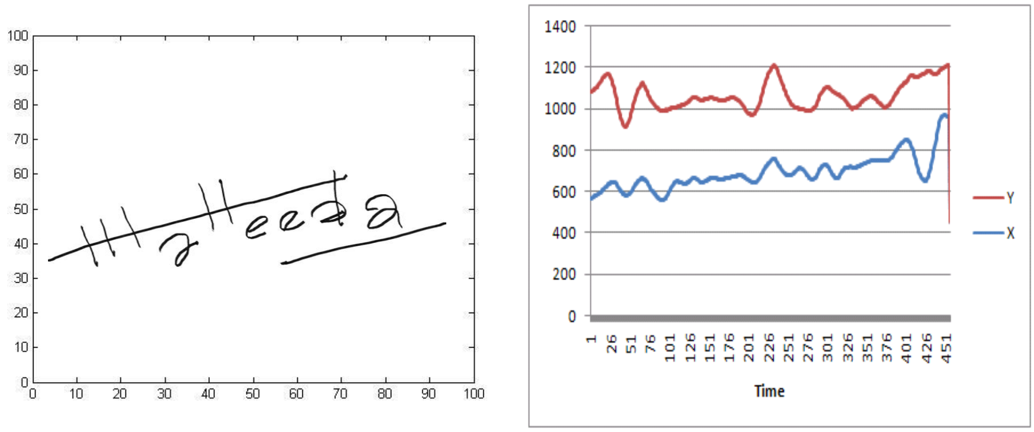


Figure 3.4 Example of a signature and the function-based representation from the gathered signature database

The signature database was captured using the above mentioned signature pad with the volunteer being seated in a comfortable position with good lighting. The volunteers were orally asked to provide their signature sample in their own time. The signers generally provided ten samples in one sitting, with this operation being repeated on two or three separate occasions resulting in twenty genuine signatures from each volunteer.

3.2.2 Preprocessing

Unlike the offline signature systems, the online systems do not suffer from noise, as a result of the scanning hardware or paper background. Nevertheless, the captured online signature signals typically have different dynamic ranges. Therefore, we have adapted a simple approach to minimize this range with respect to the maximum and minimum values [106]. A number of studies have evaluated the performance of different normalization techniques [76,90,171]. In this Thesis, we use min-max normalization approach which is expected to work well if the bounds of the distribution are known [44]. In this case, this technique shifts the minimum and maximum scores to a range between 0 and 1, respectively. Therefore, this normalization does not change the underlying distribution of the data except for a scaling factor. This is performed as shown in the following equations.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad , \quad y' = \frac{y - \min(y)}{\max(y) - \min(y)} \qquad 3.1$$

In the above equations, x and y are the initial profiles, min and max are the minimum and maximum values of x and y original profiles respectively, while x' and y' are the transformed profiles.

3.2.3 Global Feature Extraction

The discriminative power of the features in the reference set plays a major role in the entire identification process, as it is important to find features that are invariant with respect to slight changes in intra-class signatures, yet powerful enough to be used to discriminate other signature's classes. Large number of features has been reported which can be broadly classified as local and global. Global features refer to the parameters extracted from a complete signature signal, such as average writing speed, total signing duration, number of pen-ups, number of strokes and standard deviation of the velocity and acceleration. Whereas, local features analyse signatures based on specific sampling points, such as the slope of the tangent at each point, velocity, the centre of mass and average speed within a stroke. Some approaches combine both global and local features to improve the overall accuracy [46,142]. An analysis of the feasibility and consistency of different features for signature verification is thoroughly investigated in [106].

Primarily, our interest is to find the most reliable and suitable set of dynamic features to be used in our approach, so we decided to consider global features for many reasons. Such features for one reason are simple to compute with a minimum preprocessing effort to be performed on raw data, and there is no need to maintain the original signatures once the features are extracted. In fact, using only small number of such features achieved approximately 89% accuracy [88].

Table 3.1 lists the 31 global features that we have used in this study. They represent a collection of some of the statistical features that have been widely used, studied, and reported in literature [46,96,106].

Table 3.1 Implemented features

Feature identifier		Description
1.	SNx	Mean of all normalized coordinates in the X plane
2.	SNy	Mean of all normalized coordinates in the Y plane
3.	Smax	Number of times the pen was lifted over the entire signature.
4.	Svx	Mean of velocity over all coordinates in the X plane
5.	Svy	Mean of velocity over all coordinates in the Y plane
6.	Sax	Mean of acceleration over all coordinates in the X plane
7.	Say	Mean of acceleration over all coordinates in the Y plane
8.	SR	Rhythm or the speed of pen tracing out the signature[165]
9.	RMSvx	Root mean square of velocity in the X plane
10.	RMSvy	Root mean square of velocity in the Y plane
11.	RMSax	Root mean square of acceleration in the X plane
12.	RMSay	Root mean square of acceleration in the Y plane
13.	MaxAx	Maximum acceleration in the X plane
14.	MaxAy	Maximum acceleration in the Y plane
15.	MaxVx	Maximum velocity in the X plane
16.	MaxVy	Maximum velocity in the Y plane
17.	R	Correlation co-efficient
18.	Zvx	Sign changes within velocity in the X plane[106]
19.	Zvy	Sign changes within velocity in the Y plane[106]
20.	Zax	Sign changes within acceleration in the X plane[106]
21.	Zay	Sign changes within acceleration in the Y plane[106]
22.	xAz	Number of zeroes in acceleration in the X plane[106]
23.	yAz	Number of zeroes in acceleration in the Y plane[106]
24.	Savxy	Root mean square of (x,y) coordinates

25.	Npoints	Number of x,y within signature
26.	Sdvx	Standard deviation of velocity in the X plane
27.	Sdvy	Standard deviation of velocity in the Y plane
28.	Sdax	Standard deviation of acceleration in the X plane
29.	Sday	Standard deviation of acceleration in the Y plane
30.	Dx	Sum of changes between each consecutive points within X-coordinate (signature path horizontal length: total displacement in the X plane)
31.	Dy	Sum of changes between each consecutive points within Y-coordinate (signature path vertical length: total displacement in the Y plane)

The following is an explanation of some of the features used in this chapter

Mean velocity in the X plane Svx =

$$\frac{1}{N}\sum_{i=1}^{N-1} \left((x_{i+1} - x_i) \ / (t_{i+1} - t_i) \right) \qquad 3.2$$

Mean velocity in the Y plane Svy =

$$\frac{1}{N}\sum_{i=1}^{N-1} \left((y_{i+1} - y_i) \ / (t_{i+1} - t_i) \right) \qquad 3.3$$

Mean acceleration in the X plane Sax=

$$\frac{1}{N}\sum_{i=1}^{N-1} \left((Vx_{i+1} - Vx_i) \ / (t_{i+1} - t_i) \right) \qquad 3.4$$

Mean acceleration in the Y plane Say=

$$\frac{1}{N}\sum_{i=1}^{N-1} \left((Vy_{i+1} - Vy_i) \ / (t_{i+1} - t_i) \right) \qquad 3.5$$

Mean Rhythm SR=

$$\frac{1}{N}\sum_{i=1}^{N-1} \left((x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2 \right)^{0.5} \qquad 3.6$$

Root Mean Square of velocity in the X plane RMSvx =

$$\left(\frac{1}{N}\sum_{i=1}^{N} (Vx_i)^2 \right)^{0.5} \qquad 3.7$$

Root Mean Square of velocity in the Y plane RMSvy =

$$\left(\frac{1}{N}\sum_{i=1}^{N} (Vy_i)^2 \right)^{0.5}, i = 1, \ldots, N \qquad 3.8$$

Root Mean Square of acceleration in X plane RMSax =

$$\left(\frac{1}{N}\sum_{i=1}^{N}(Ax_i)^2\right)^{0.5}, i = 1,\dots,N \tag{3.9}$$

Root Mean Square of acceleration in the Y plane RMSay =

$$\left(\frac{1}{N}\sum_{i=1}^{N}(Ay_i)^2\right)^{0.5}, i = 1,\dots,N \tag{3.10}$$

Correlation co-efficient R=

$$\frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{n\sum x_i^2 - \left(\sum x_i\right)^2}\sqrt{n\sum y_i^2 - \left(\sum y_i\right)^2}} \tag{3.11}$$

Standard deviation of velocity in the X plane Sdvx =

$$\left(\left(\sum_{i=1}^{N}\left(Vx_i - \overline{X}\right)^2\right)/(N-1)\right)^{0.5} \tag{3.12}$$

Where $\overline{X} = \frac{1}{N}\sum_{i=1}^{N} x_i$ 3.13

Standard deviation of velocity in the Y plane Sdvy =

$$\left(\left(\sum_{i=1}^{N}\left(Vy_i - \overline{X}\right)^2\right)/(N-1)\right)^{0.5} \tag{3.14}$$

Standard deviation of acceleration in the X plane Sdax =

$$\left(\left(\sum_{i=1}^{N}\left(Ax_i - \overline{X}\right)^2\right)/(N-1)\right)^{0.5} \tag{3.15}$$

Standard deviation of acceleration in X plane Sday =

$$\left(\left(\sum_{i=1}^{N}\left(Ay_i - \overline{X}\right)^2\right)/(N-1)\right)^{0.5} \tag{3.16}$$

Sum of changes between each consecutive points within X-coordinate Dx = $\sum_{i=1}^{N-1}(x_{i+1} - x_i)$ 3.17

Sum of changes between each consecutive points within Y-coordinate Dy = $\sum_{i=1}^{N-1}(y_{i+1} - y_i)$ 3.18

3.2.4 Classification

The problem of signature authentication is considered a two-class pattern recognition mission, where the signature sample is classified either genuine or not. Unfortunately signatures tend to vary slightly each time they are captured. In addition, variations in acquired samples and template data make signature verification a challenging pattern recognition problem.

To deal with the problem of online signature authentication, researchers have investigated a variety of techniques which include DTW [86], signal correlation [126], neural networks such as MLP [108], time-delay neural networks [20], HMM [37,188], Euclidean and other distance measure approaches[74]. Other classification paradigms such as k-nearest neighbor (k-NN) methods or support vector machine (SVM) are also investigated in [2] and [55] respectively.

Throughout this thesis we considered a number of classification algorithms to evaluate the benefits of the integrated behavioural/psychosocial biometrics. These algorithms are quite popular and well known in pattern recognition literature. Yet, with few notable exceptions, their usefulness for biometric recognition is still to be evaluated. Among the wide diversity of classifiers, we selected the Naïve Bayes classifier and the k-NN algorithm for comparison as they are both distinguished, clear, and they both perform well in many classification problems.

3.2.4.1 Naïve Bayes Classifier

The Bayesian method is one of the most popular machine learning methods. Bayesian networks are now an increasingly powerful tool for reasoning under uncertainty, supported by a wide range of mature academic and commercial software tools. They are now being applied in many domains, including environmental and ecological modelling, bioinformatics, medical decision support, many types of engineering, robotics, military, financial and economic modelling, education, forensics, emergency response, surveillance, and so on [123].
A simple form of Bayes networks is Naïve Bayes classifier [50,107]. Naïve Bayes classifier also known for his inherent robustness to noise is characterized by the assumption that the feature attributes are independent of one another given the class and all the probability estimations from the training sample are accurate. The naïve Bayes have a history as a successful classifier in text classification [78,120].

Naïve Bayes classifiers classifier uses the Bayes theorem to predict the category for each unseen instance. Naïve Bayes classifiers operate on data sets where each example x consists of training samples made up from discrete-valued attributes (a_1, a_2 ,..., a_i) and the target function *f(x)* can take on any value from a pre-defined finite possible classes set V=(v_1, v_2 ,..., v_j). For example, let x be a description of the day's weather conditions (e.g. sunny, windy or rainy), and let V be a set of activities (e.g. play golf, walk, stay at home). If the task is to predict the day's activity on the basis of the weather condition, f(x) would be a mapping from x to V. Classifying test instances involves calculating the most probable target value v_{max} which is defined as

$$v_{\max} = \max_{v_j \in V} P(v_j | \; a_1, a_2, \ldots, a_i) \qquad 3.19$$

With the use of Bayes theorem, v_{max} can be rewritten as [123]

$$v_{max} = \frac{\max_{v_j \in V} P(a_1, a_2, ..., a_i | v_j) P(v_j)}{P(a_1, a_2, ..., a_i)} \quad 3.20$$

$$v_{max} = \max_{v_j \in V} P(a_1, a_2, ..., a_i | v_j) P(v_j) \quad 3.21$$

With assuming that features values are class-conditional independence, given the target class, the formula 3.21 can be rewritten as

$$v_{max} = \max_{v_j \in V} P(v_j) \prod_i P(a_i | v_j) \quad 3.22$$

Where V is the target output of the classifier and $P(a_i|v_j)$ and $P(v_i)$ can be calculated based on their frequency in the training instances. Thus, Naïve Bayes assigns a probability to every possible value in the target range. The resulting distribution is then condensed into a single prediction. Further details on the Naive Bayes can be found in [122].

3.2.4.2 k-Nearest Neighbor (k-NN) Classifier

The k Nearest Neighbors classifier (k-NN) is an instance-based learning algorithm which has been studied in pattern recognition, data analysis, and data mining problems for a long time [166]. The k-NN is a supervised machine learning algorithm which is used for classification based on the closest training samples in the feature space. For the purpose of identification the training objects are represented as vectors in a multidimensional feature space, each with a class label. In this method, an unknown object is assigned to the class that is most frequent among its k nearest neighbours.

According to the algorithm, the value of *k* should be a positive integer between 1 and the total number of observations. If *k = 1,* then the algorithm assigns the objects to the class of its nearest neighbour. The k-NN algorithm is commonly based on measuring the distance test sample and the specified training samples using the Euclidean distance discriminant function [29]. The Euclidean distance metric has the following form

$$G(T) = (1/n) \sum_{i=1}^{n} \left(\frac{\overline{X}_i - \sigma_i}{s_i} \right)^2 \quad 3.23$$

where, as defined earlier, T is the test signature and $\overline{X}_i$ and σ_i are, respectively, the ith feature's reference *mean* and reference *standard deviation*. Other kinds of distance measures like the Manhattan distance could be used also.

In all the experiments throughout this thesis, we have used the Euclidean distance based k-NN classifier. The advantage of this classifier is its conceptual simplicity and the fact that it does not require any training. The basic idea of k-NN algorithm is presented in Figure 3.5.

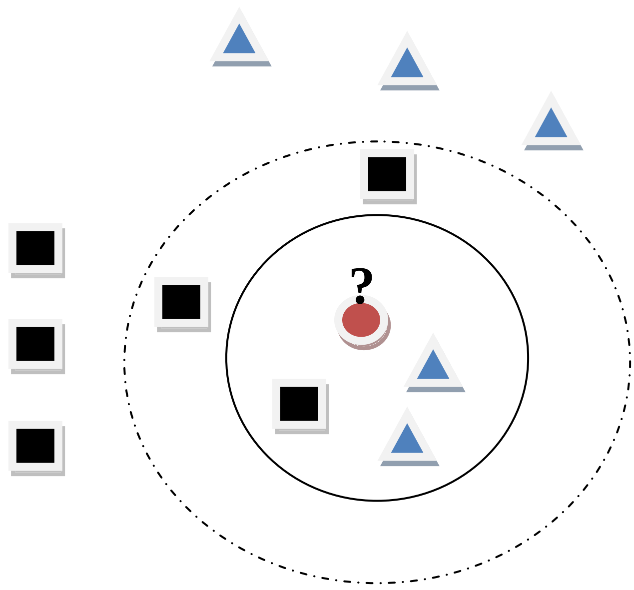


Figure 3.5 An illustration of k-NN technique.

The unknown sample (the circle) could be classified either to the first class of squares or to the second class of triangles based on the value of k. If k = 3 it will be assigned to the triangle class as there are 2 triangles and only 1 square inside the inner circle. But if k = 5, then it will be assigned to the square class as there are 3 squares and 2 triangles inside the outer circle.

3.2.5 Feature Selection Techniques

Feature selection, also known as variable selection, feature reduction, attribute selection or variable subset selection, is critical in designing a biometric based recognition system. Feature selection techniques help in recognizing and eliminating much of the irrelevant and redundant features. Consequently, it reduces the dimension of feature space, which is important for the success of online implementation in biometric recognition. Moreover, researchers have shown [3,101,102,124,148] that irrelevant and redundant training features can negatively effects the classifier performance.

A number of feature selection algorithms can be applied to perform feature selection of biometrics features. In this chapter, we will study the effect of three feature selection techniques: Rough set, PCA and correlation-based feature selection on online signature identification.

3.2.5.1 Rough Set Based Feature Reduction Technique

Rough sets theory was first introduced by the Polish computer scientist Zdzisław I. Pawlak in the 1980's as a new mathematical tool to handle uncertainty, imprecision and vagueness of decision system [139]. Since then, large number of researchers contributed to the further development of the field by extending and applying the theory. It is based on the concept of approximation spaces and models of the sets and concepts. Due to the fact that the rough set had shown ability to extract dependency rules directly from data itself and it do not require any preliminary or further information about data, this theory has been applied successfully in many domains. The two main applications of the classical Rough Sets theory are in feature reduction and classification. It was later applied within other areas such as unsupervised learning.

First we will present here some preliminaries of rough set theory, which are relevant to this chapter. For details we refer the reader to [94,113,140].

1. Rough set theory preliminaries

In rough sets theory, the data is described as a table, called a decision table or some references information system. Rows of a decision table correspond to objects (*observations*), and columns correspond to features (*attributes*). In an information system, every object of the universe is associated with a set of features to describe it. Objects characterized by the same information are indiscernible in consideration of the available information about them. Any set of indiscernible objects is called elementary sets (*neighborhood*). Any union of elementary sets is called a crisp (*precise*) set; and any other set is referred to as rough (imprecise, vague).

The rough set approach is characterized by its lower and upper approximations to handle the inconsistent information. The lower approximation consists of all objects which surely belong to the subset of interest whereas the upper approximation contains all objects which possibly belong to the subset. The difference between the upper and the lower approximation constitutes the boundary region.

Definition 1 (Information System, [94])

An information system can be represented as

$$S = (U, \Omega, V_a, f_a) \qquad 3.24$$

Where U — a finite , nonempty, closed set of objects(observations, examples) called the universe; Ω — a nonempty, finite set of features (attributes); $\Omega = C \cup D$ in which C is a finite set of condition features and D is a finite set of decision features; For each $a \in \Omega$ is called the domain of a; f_a — an information function $f_a : U \rightarrow V_a$.

Definition 2 (Indiscernibility Relation , [94])

Indiscernibility, which refers to the similarities among different objects, is a main concept in rough set theory. Every subset of features $B \subseteq A$ induces indiscernibility relation

$$Ind_B = \{(x, y) \in UxU : \forall_{a \in B} a(x) = a(y)\}. \quad 3.25$$

For every x $\in$ U, where there is an equivalence class $[x]_B$ in the partition of U defined by Ind_B. The indiscernibility relation is an equivalence relation and it splits the objects into a family of equivalence classes, called elementary sets.

Definition 3 (Lower and Upper Approximation, [94])

In the rough sets theory, the approximation of sets is introduced to deal with inconsistency. A rough set approximates traditional sets using a pair of sets named the lower and upper approximation of the set. Given a set $B \subseteq A$, the lower approximations of a set $Y \subseteq U$ are defined as

$$\underline{B}X = \{x \mid [x]_B \subseteq X\} \quad 3.26$$

or the set of all elements of U which can be with certainty classified as elements of *X*.

The upper approximation of X with respect to B is defined as

$$\overline{B}X = \{x \mid [x]_B \cap X \neq \phi\} \quad 3.27$$

or all objects whose equivalence classes have a nonempty intersection with X. In other word, it contains all objects which can possibly be classified as belonging to the set X. Figure 3.6 shows an example where the indiscernibility relation partitions the domain into grids, the semi oval shape is the set to be approximated, the dark grey area is the lower approximation and the light grey area is the upper approximation.

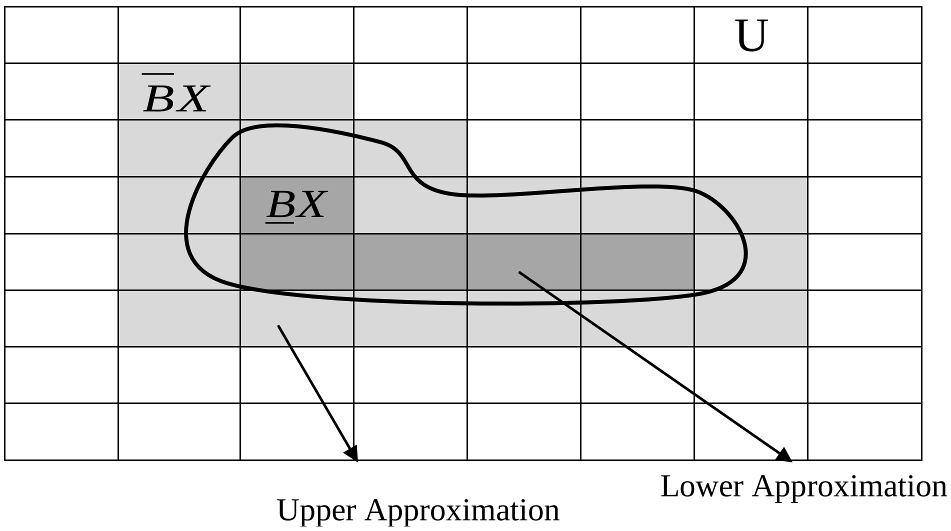


Figure 3.6 Representation of the set approximations

Definition 4 (Lower Approximation and Positive Region, [94])

The positive region $POS_C(D)$ is defined by

$$POS_C(D) = \bigcup_{X:X \in U/Ind_D} \underline{C}X \qquad 3.28$$

$POS_C(D)$ is the set of all objects in U that can be uniquely classified by elementary sets in the partition U/Ind_D by means of C [94].

Definition 5 (Upper Approximation and Negative Region, [94])

The negative region $NEG_C(D)$ is defined by

$$NEG_C(D) = U - \bigcup_{X:X \in U/Ind_D} \overline{C}X \ , \qquad 3.29$$

that is the set of all objects can be definitely ruled out as member of X.

Definition 6 (Boundary Region, [74])

The distance between upper and lower approximations of a set X constitutes the boundary region that consists of equivalence classes having one or more elements in common with X. The boundary region it is defined by the following formula

$$BND_B(X) = \underline{B}X - \overline{B}X \qquad 3.30$$

A approximation of a rough set can be described using the accuracy of the approximation which is measured as the ratio of the lower and the upper approximations.

$$\alpha_B(X) = \left|\underline{B}X\right| / \left|\overline{B}X\right| \qquad 3.31$$

where $|\bullet|$ denotes the cardinality of $X \neq \phi$. X is definable with respect to B if $\alpha_B(X) = 1$, otherwise X is *rough* with respect to B.

2. Reduct and Core

Definition 7 (Degree of Dependency, [94])

The degree of dependency between D and C can be defined as:

$$\gamma(C, D) = \left|POS_C(D)\right| / \left|U\right| \qquad 3.32$$

Definition 8 (Reduct)

Given a classification task mapping a set of features C to a set of class labels D, a reduct set is defined with respect to the power set *P(C)* as the set $R \subseteq C$ such that:

$$R = \{A \in P(C) : \gamma(C,D) = \gamma(R,D)\} \quad 3.33$$

That is, the reduct set is the set of all possible reducts of the equivalence relation denoted by C and D.

3. Significance of the Attribute

Significance of features expresses the importance of the features by assigning a real number from the closed interval [0, 1] to it.

Definition 9 *(Significance,* [94]***).***

For any feature $a \in C$, we define its significance ξ with respect to D as follows:

$$\xi(a,C,D) = \frac{|POS_{C \setminus \{a\}}(D)|}{|POS_C(D)|} \quad 3.34$$

A reducted feature set can be found by designing a heuristic attribute reduction algorithm through selecting the attributes with the maximum significance interactively based on the significance of a feature [110].

4. Decision Rules

In the perspective of supervised machine learning, an important task is to discover the rules of classification from the instances provided in the decision tables. The decision rules capture hidden patterns and predict the class of unseen objects. Rules represent the extracted knowledge which can be used when classifying unseen objects and the dependencies in the dataset. Whenever the reduct is found, the task of creating specific rules for the value of the decision feature of the information system is practically completed.

To convert a reduct into a rule, one only has to bind the condition feature values of the object class from which the reduct originated to the corresponding features of the reduct. Afterwards, to complete the rule, a decision part comprising the resulting part of the rule is added. Rules generated from a training set will be used to classify unseen objects.

5. Rough Sets Data Analysis Techniques

In this section, we discuss in detail the proposed rough set scheme to analyse online signatures which consists of two stages. These stages include data discretization and attribute reduction.

Stage 1: Data discretization

Data discretization , also referred to as discretization in machine learning, significantly improves the performance of data mining algorithm by converting the original continuous input space into finite set of intervals with least loss of information. It is a familiar data transformation procedure that involves finding the discretization intervals or the cut-off points in the data sets which divide the data into intervals. After that, it maps the whole values lying within an interval to the same value. Data discretization concept will lead to reducing the size of the attributes value set. In this chapter we adopt the rough sets with boolean reasoning (RSBR) algorithm proposed by Zhong et al. [193] for the discretization of continuous-valued attributes. The main advantage of RSBR is that it combines discretization of real valued attributes and classification.

Stage 2: Attribute reduction

We apply a dynamic reduct technique to integrate a decision rule from decision table. The process of computing dynamic reduct can be seen as a combining normal reduct computation with re-sampling technique. Simply the idea consists of three steps. The first step is randomly sampling a family of subsystems from the universe. In the second step, computes the reduction of each sample. The final step is to keep the reduct that occur most frequently as it is the most stable one.

3.2.5.2 Feature selection using PCA

Principal component analysis (PCA), also known as Karhunen–Love transform, is one of the most widely used linear dimensionality reduction algorithm. PCA was introduced for the first time by Karl Pearson (1901). PCA is a linear transformation applied to a set of observations in order to obtain a new orthogonal coordinate system. This transformation is defined in such a way that the first axis lies along the direction of greatest variance in the data set, and each succeeding component in turn lies along the direction of the second greatest variance, and so on. These new axes are known as the *principal components*.

PCA has been successfully used as an initial step in many pattern recognition applications by first obtaining the principal components and then discarding the dimensions contributing the least to the variance of the data set. The formulation of standard linear PCA, mapping the original N-dimensional biometric feature space into an M-dimensional feature space where m < n, is as follows. Let us consider a set of N feature vectors X_n, n = 1, ... , N, PCA finds a linear transformation W^T mapping the original N−dimensional feature space into an M−dimensional feature space. Denoting by $W \in R^{nxm}$ a matrix with orthonormal columns, the new feature vectors coordinates y_k are defined by the following linear transformation:

$$y_k = W^T x_K, \qquad k = 1,2,..., N \tag{3.35}$$

3.2.5.3 Correlation-based Feature Selection

In this paper, we applied the correlation-based objective selection (CFS) algorithm which has been shown [58] to be quite successful in feature evaluation and selection. The CFS algorithm evaluates the importance of feature sets on the basis of the following hypothesis: "*A good feature subsets is one that contains features highly correlated with the classification, yet uncorrelated to each other*" [59].

The following objective function, also known as Pearson's correlation coefficient, gives the merit of a feature subset consisting of *k* features [59]

$$r_{zc} = \frac{k\overline{r_{zi}}}{\sqrt{k + k(k-1)\overline{r_{ij}}}} \qquad 3.36$$

where $\overline{r_{zi}}$ is the is average feature value of all feature-classification correlations and $\overline{r_{ij}}$ is average value of all feature-feature correlations. The CFS based feature selection algorithm uses r_{zc} to search the feature subsets using the best first search [60].

The CFS algorithm starts the search with evaluating of all the individual features as a separate subset. The algorithm retains the feature subset with the highest objective function. Afterwards, the feature subset space can be enlarged by adding all possible combinations of new features to the resulting combinations. The search process returns to the next unexpended subset if the new added feature does not show any improvement in its accuracy. The search will be aborted if the addition of new features to the subset does not show significant improvement.

3.3 Database and Experiments

This section describes the experimental setup, including database and the assessment protocol that we have built in order to evaluate the proposed authentication schemes.

3.3.1 Database

The proposed method has been started by building our own database to form the nucleus of a local database. It contains 2160 signatures gathered from108 different volunteer subjects. Among those subjects, 60 are females and two are left-handed. Each subject was asked to contribute 20 signatures collected in two sessions that were held two to four weeks apart. Ten signatures were collected from each subject during each session. An example of signatures of some volunteers who contributed to build the signature database is given in Figure 3.6. There were no constraints on how to sign, so the subjects signed in their most natural way; in an arbitrary orientation. Therefore, there was a significant intra-class deformation and variation among signatures that belong to the same subject. Figure 3.7 depicts corresponding x and y profiles of two signatures from the same subject captured during different sessions.

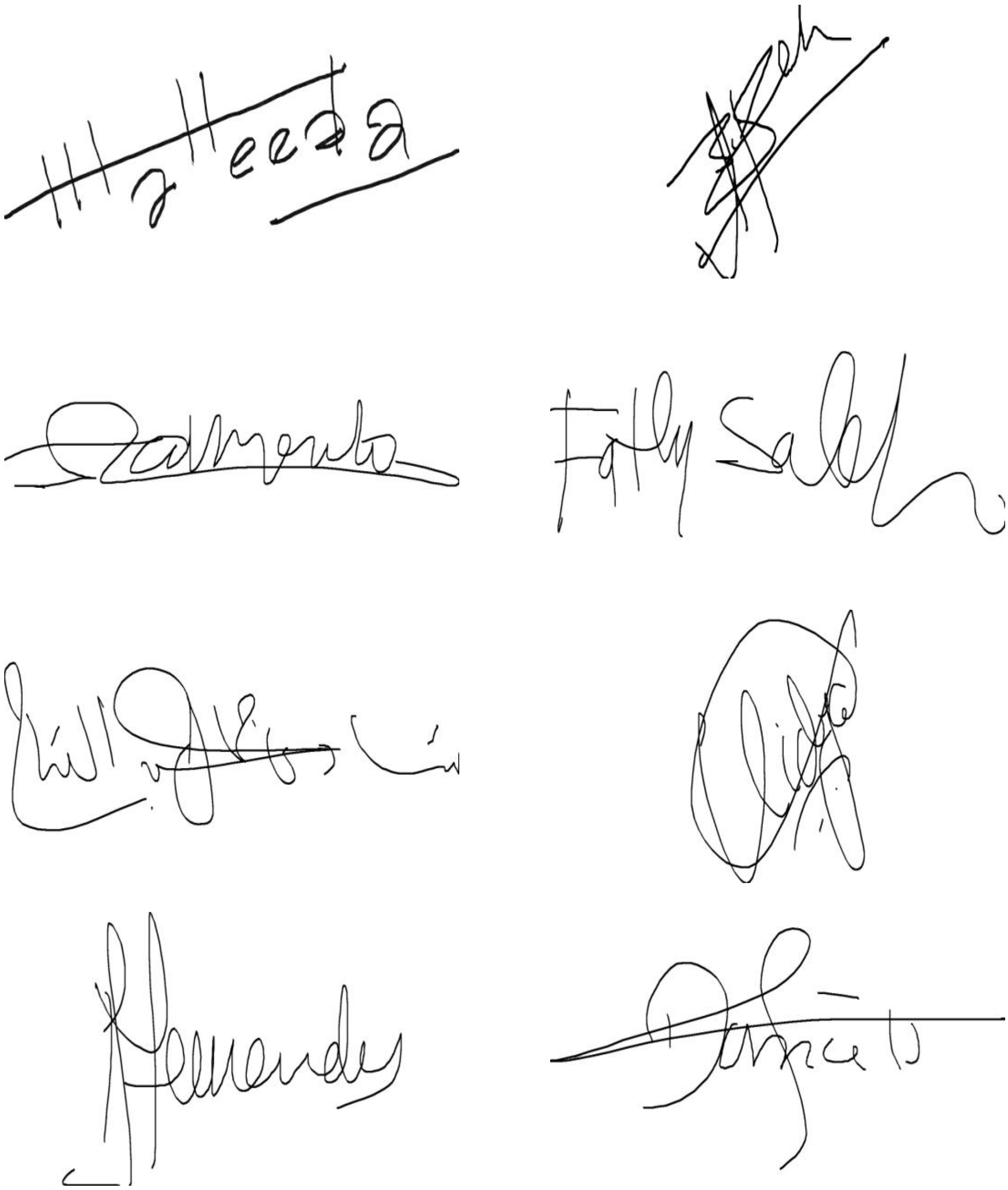


Figure 3.7 Example of some signatures of volunteers who contributed to the signature database

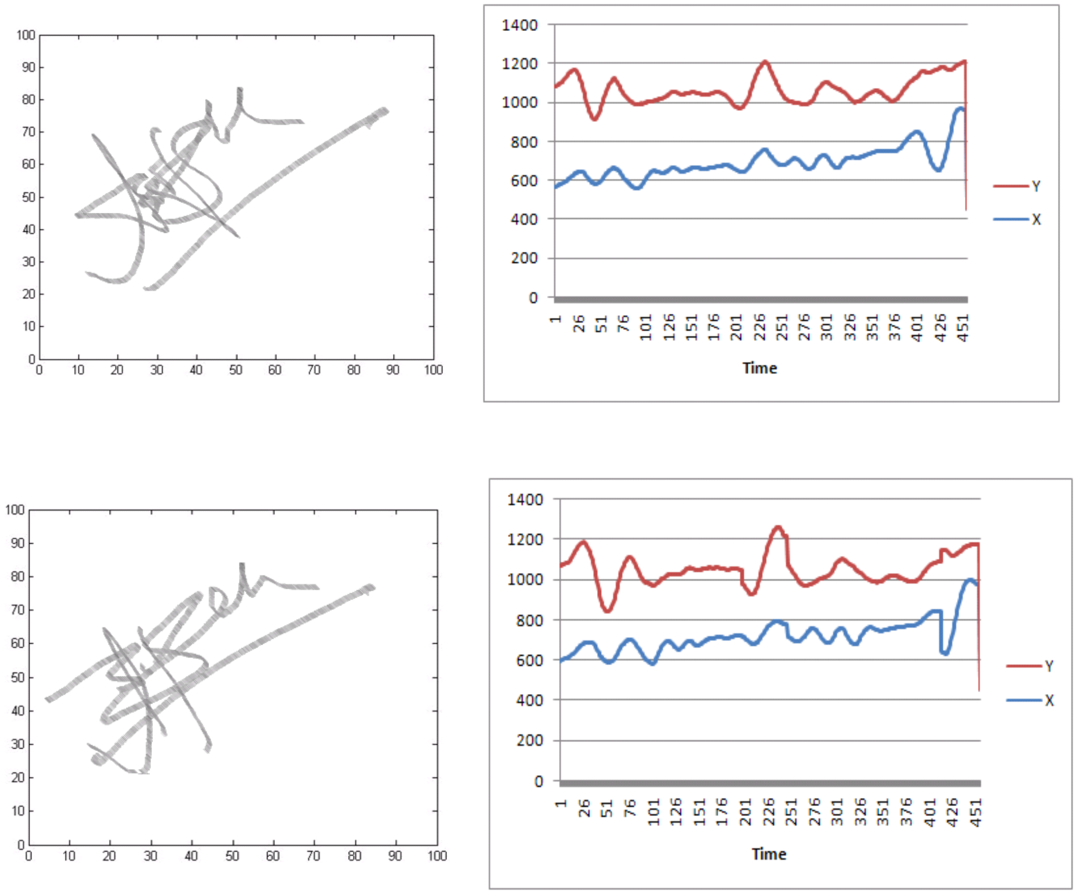


Figure 3.8 Two sample signatures from the same volunteer and their corresponding (x,y) coordinate profiles from the collected database

3.3.2 Experiments

A number of experiments were conducted to evaluate the classifiers as well as the discriminative potential of the feature sets. For all of the experiments throughout the thesis we implemented 10-fold cross validation.

K-fold Cross-Validation

Cross-validation is a method designed for estimating the generalization error based on "resampling" [162]. Cross-validation technique allows using the whole data set for training and testing. In k-fold cross-validation procedure, the relevant dataset is partitioned randomly into approximately equal size k parts called folds and trained k times, each time leaving out one of the folds from training process, whilst using only the omitted fold to compute error criterion. Then the average error across all k trials is estimated as the mean error rate and defined as

$$E = \frac{1}{k}\sum_{i=1}^{k} e_i \qquad 3.37$$

where, e_i is error rate of each k experiment. Figure 3.10 depicts the concept behind k-fold cross validation.

k= 1	k=2	k=3	…	k=K
Train	Train	Validate		Train

Figure 3.9 Data partitioning using k-fold cross-validation.

The whole dataset is divided into K folds. One fold (k = 3, in this example) is set aside to validate the data of testing and the remaining K − 1 folds are used for training. The entire procedure is repeated for each of the K folds.

A number of studies found that the value of 10 for k leads to adequate and accurate classification results [57]. Therefore, we have used k =10 folds for training and the remaining k-1 folds for testing in all the experiments conducted in this work.

The first set of experiments was conducted using the entire database which contains 2160 signatures.

- **Experiments 1** and **2:** In these experiments, all 31 features shown in Table 3.1 were used with the Naïve Bayes and k-NN classifiers. The correct classification rate achieved by the Naïve Bayes was **97.1%**. Whereas, the k-NN classifier achieved **98.33%**.

- **Experiment 3**: In this experiment, the rough set approach was used to find the minimal reduct set (Definitions 7 and 8) of features. This has resulted in the following 9 features: {1,2,3,4,5,8,10,11,30} from Table 3.1 which corresponds to the following set of features: {SNx, SNy, sMax, SVx, SVy, SR, RMSVy, RMSAx, Dx}. Using this features set with the Naïve Bayes classifier resulted in classification accuracy of **96.3%**. Table 3.2 shows some statistics of the above minimal reduct set.

- **Experiment 4:** In this experiment, the k-NN classifier was used with the Rough Set minimal reduct set comprised of nine features. The classification accuracy achieved was **95.41 %**.

- **Experiments 5 and 6:** In these experiments, the PCA was applied prior to classification to reduce the number of feature from 31 to 14. The classification accuracy achieved by the k-NN classifier was **90.83%** and Naïve Bayes classifier achieved **88.19**%.

Experiments 7 and 8: Applying the CFS algorithm has revealed 13 redundant and irrelevant features. This has resulted in the following 18 features: {SNx, SNy, St, SVy, SR, MaxxA, MaxyA, MaxyV, r, ZVx, ZVy, xAz, SAvxy, Npoints,Dx,Dy} from Table 3.1. The performance of 18 relevant online signature features, or feature subset, is also illustrated in Table 3.3. Using the 18 features set with the Naïve Bayes classifier resulted in classification accuracy of **98.01**%.

Table 3.2. Statistics of minimal reduct set

Feature	***Mean***	***Standard Deviation***	***Correlation***
1	0.52	0.077	0.174
2	0.512	0.086	0.117
4	4.176	2.67	0.088
5	-0.0004	00.000327	-0.224
6	0.000297	-0.00035	-0.095
9	0.719	0.386	-0.008
11	0.011	0.004	0.181
12	0.002	0.001	-0.087
30	2520.437	1621.87	0.044

Table 3.3 shows the summary of the results of the eight experiments carried above. It clearly demonstrates the suitability and superiority of using the proposed Rough Set approach for feature reduction in online signature identification. The Rough Set has achieved classification accuracy of **97.26**% with 9 features only.

Table 3.3 Summary of the entire database classification results

Experiment	***Total number of signatures***	***Number of Features***	***Feature Reduction***	***Classifier***	***Accuracy rate %***
1	2160	31	-	Naïve Bayes	97.5
2	2160	31	-	k-NN	98.33
3	2160	9	Rough set	Naïve Bayes	97.26
4	2160	9	Rough set	k-NN	95.41
5	2160	15	PCA	Naïve Bayes	88.19
6	2160	15	PCA	k-NN	90.83
7	2160	14	CFS	Naïve Bayes	98.01
8	2160	14	CFS	k-NN	98.14

It can be observed from the Table 3.3 that the best performance for online signature recognition is achieved with k-NN classifier before applying any feature reduction technique. Overall, the performance of k-NN is better than naïve Bayes. The evaluation of 32 online signature features from the training set, using the Rough Set, PCA and CFS algorithms has revealed a number of redundant and irrelevant features. The performance of PCA has been the worst and this may be due to the large number of features that make the repeated portioning of data difficult. The performance of CFS is better than the Rough Set. However, the performance of Rough Set is promising with considerably less number of features.

As we mentioned earlier, the signatures were collected over 2 sessions, two to four weeks apart. In each session, 10 signatures were collected from each of the 108 subjects. In the following 2 experiments we test each session separately. The next set of experiments is carried over the dataset gathered during the first enrolment session, which contains of 1080 signatures.

- **Experiments 9** and **10:** in these experiments, all 31 features shown in Table 3.4 were used with the Naïve Bayes and k-NN classifiers. The best classification rate achieved by the k-NN classifier was **98.51%**, whereas the Naïve Bayes achieved **96.75%**.

- **Experiment 11**: In this experiment, the Naïve Bayes classifier resulted in classification accuracy of **95.18%** with the Rough Set minimal reduct set.

- **Experiment 12:** In this experiment, the k-NN classifier was used with the Rough Set minimal reduct set comprised of nine features. The classification accuracy achieved was **95.83%**.

- **Experiments 13 and 14:** In these experiments, the PCA was applied to reduce the number of feature is from 31 to 14. The classification accuracy achieved by the k-NN classifier was **90.18%** and Naïve Bayes classifier achieved **80.64**%.

- **Experiments 15 and 16:** In these experiments, the CFS was applied to reduce the number of feature is from 31 to 18. The classification accuracy achieved by the k-NN classifier was **98.05%** and Naïve Bayes classifier achieved **96.48**%.

Table 3.4 Summary of first session dataset classification results

Experiment	***Feature Reduction***	***Number of Features***	***Classifier***	***Accuracy rate%***
9	-	31	Naïve Bayes	96.75
10	-	31	k-NN	98.51
11	Rough	9	Naïve Bayes	95.18
12	Rough	9	k-NN	95.83
13	PCA	15	Naïve Bayes	80.64
14	PCA	15	k-NN	90.18
15	CFS	14	Naïve Bayes	96.48
16	CFS	14	k-NN	98.05

The next set of experiments is carried over the dataset gathered during the second enrolment session.

- **Experiments 17** and **18:** In these experiments, all the 31 features shown in Table 3.1 were used with the Naïve Bayes and k-NN classifiers without applying any feature reduction technique. The correct classification rate achieved by the Naïve Bayes was **96.2%**. Whereas, the k-NN classifier achieved **97.41%**.

- **Experiment 19**: In this experiment, the Naïve Bayes classifier resulted in classification accuracy of **95.92%** with the minimal reduct set.

- **Experiment 20:** In this experiment, the k-NN classifier was used with the minimal reduct set comprised of nine features. The classification accuracy achieved was **95%**.

- **Experiments 21 and 22:** In these experiments, the two classifiers were trained and tested with the PCA-based reducted feature set. The classification accuracy achieved by the k-NN classifier was **89.53%** and Naïve Bayes classifier achieved **95**%.

- **Experiments 23 and 24:** In these experiments, the two classifiers were trained and tested with the CFS-based reducted feature set. The classification accuracy achieved by the k-NN classifier was **97.4%** and Naïve Bayes classifier achieved **96.01**%.

Table 3.5 Summary of second session dataset classification results

Experiment	***Feature Reduction***	***Number of Features***	***Classifier***	***Accuracy rate%***
17	-	31	Naïve Bayes	96.20
18	-	31	k-NN	97.41
19	Rough	9	Naïve Bayes	95.92
20	Rough	9	k-NN	95.00
21	PCA	15	Naïve Bayes	95.00
22	PCA	15	k-NN	89.53
23	CFS	14	Naïve Bayes	96.01
24	CFS	14	k-NN	97.40

Tables 3.4 and 3.5 summarize the experimental results for the online signature identification in the two acquisition sessions. It can be seen that the best performance for online signature recognition is achieved with k-NN classifier before applying any feature reduction technique. Overall, the performance of k-NN is better than Naïve Bayes. The reason behind this performance drop is that the Naïve Bayes classifier is negatively affected by redundant attributes as a result of its primary assumption that all the attributes are conditionally independent [78,120].

The difference in results between the two signing sessions was expected. The subjects themselves were less enthusiastic in completing the second signing session when they were approached few weeks later. This has resulted in higher intra-variations in the signatures than those of the first session.

The purpose of the feature reduction is to identify the significant features and eliminate the irrelevant of dispensable features to the learning task. The benefits of feature reduction are twofold: firstly, it considerably decreased the computation time. Secondly, it increases the accuracy of the resulting classification. Rough sets have been employed here to remove redundant conditional attributes from discrete-valued datasets, while retaining their information content. This approach has been applied to aid classification of online signatures, with very promising results. The analysis of experimental results in Tables 3.3-3.5 suggests that the Rough Set-based feature subset selection is capable of effectively selecting the relevant online signature features more than PCA and CFS techniques.

3.4 Summary

Handwritten signature authentication is the process of verifying the identity of a person based on his/her handwritten signature sample. A novel online signature identification scheme based on global features is proposed. The information is extracted as time functions of various dynamic properties of the signatures. A database of 2160 signatures from 108 subjects was built. Thirty-one features were identified and extracted from each signature.

Different feature reduction approaches and classifiers were applied to assess their suitability for this application The results presented in this chapter have demonstrated the success of using the proposed Rough set approach in feature reduction of online signatures. This resulted in a minimal set of nine features. The reported results from several experiments demonstrate the suitability and effectiveness of the Rough set approach in the application of online signature identification.

Chapter 4
Iris Features Extraction using Dual-Tree Complex Wavelet Transform

Biometrics-based personal authentication systems have recently gained intensive research interest due to the unreliability and inconvenience of traditional authentication systems. Biometrics recently became a vital element of any successful person identification solutions as biometric traits cannot be stolen, shared or even forgotten [80].

Among biometric technologies, iris-based authentication systems bear more advantages than other biometric technologies do. Iris offers an excellent recognition performance when used as a biometric. Iris patterns are believed to be unique due to the complexity of the underlying the environmental and genetic processes that influence the generation of iris pattern. These factors result in extraordinary textural patterns that are unique to each eye of an individual and even distinct between twins [32].

Iris is a delicate circular diaphragm lies between the cornea and the lens of the human eye. The human iris pattern varies between different individuals. The iris is considered to be one of the most stable biometric [80,72,154], as it is believed to not alter significantly during a person's lifetime. Iris recognition is the most precise personal identification biometric.
Compared with other biometrics, such as fingerprints and face, iris-based authentication has a fairly short history of use. The idea of an automatic iris authentication procedure was conceptualized and patented by Flom and Safir in 1987 [49]. Most of the common approaches reported in the literature are based on iris code and integral-differential operators suggested by Daugman [33,34,136].

The aim of this chapter is to explore the potential of deploying dual-tree complex wavelet transform and support vector machine in iris classification. The remainder of this chapter is organized as follows. The first three sections describe some relevant background and related work. Descriptions of the proposed technique for iris image preprocessing and feature extraction are given in Sections 4.4 and 4.5. Experimental results, comparisons with other methods, and discussions are reported in Section 4.6. Finally, Section 4.7 concludes the chapter.

4.1 Iris Anatomy

Iris is the "coloured ring of tissue around the pupil through which light...enters the interior of the eye." [136] The iris is located in front of the crystalline lens, and divides the anterior aqueous into the anterior and posterior chambers. The pigmented fibrovascular tissue known as stroma characterized the iris. The iris's role is to help in regulating the amount of light that enters the eye. The iris is made up of smooth muscle fibers known as sphincter and dilator, which adjust pupil size with the purpose of controlling the amount of light passing through the pupil. The sclera - often referred to as white or white of the eye- is the outer white coat of connective tissue and blood vessels surround the iris. It together with internal fluid pressure maintains the eye shape and cares for its delicate internal components [125]. The surface of the eye is covered by a curved band of strong, clear tissue called the cornea. It is the first and most powerful lens in the human eye's optical system. The cornea is transparent window of the eye through which light passes. The transparency of the cornea is due to the fact that, unlike most tissues in the body, it does not contain any blood vessels. However, the cornea receives its nourishment from the tears and aqueous humor in the chamber behind it. The anatomy of the eye is shown in Figure 4.1.

Iris naturally has a rich, distinctive and complex pattern of crypts, furrows, arching, collarette and pigment spots [136]. Each human being iris has a distinctive texture which is believed to be determined randomly during the embryonic development of the eye [32]. They are also believed to be safely considered unique even between the left and right eye of the same person [31]. Although iris colour can change based on the levels of melanin concentration and distribution within the iris stroma, yet for most of a human's lifetime the appearance of the iris is relatively constant [141].

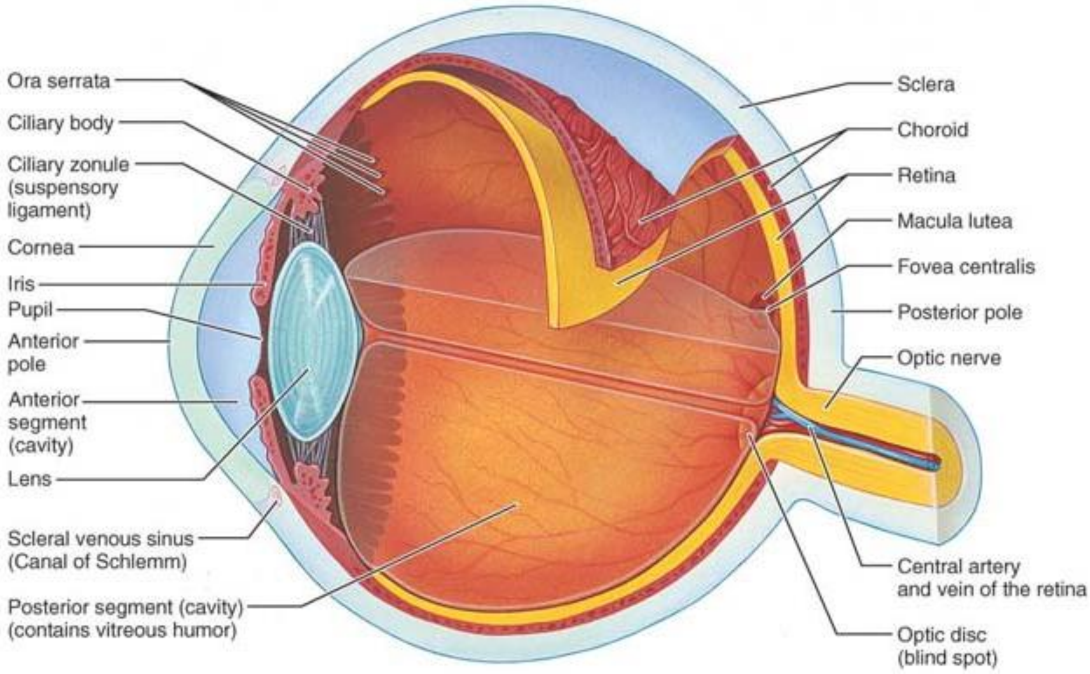


Figure 4.1 Diagrammatic view of the anatomy of the eye, *adapted from* [125]

4.2 Iris as a Biometric

Iris recognition as a reliable method for identity authentication is playing an important role in many mission-critical applications such as access control and border checkpoints for several reasons [16]:

- Iris is an internal organ of the eye, physically protected from external environment by the cornea. This makes it more consistent than fingerprints which are more susceptible to worn out due to age or manual labour.

- As the iris starts to develop in the third month of gestation, the structures creating its pattern are mainly completed by the eighth month [141]. Then it does not vary throughout one's lifetime. Furthermore, the forming of iris depends on the initial environment of embryo. Therefore, the texture patterns of the iris don’t correlate with genetic determination. Consequently irises of genetically identical twins are extremely distinct. Actually, the left and the right irises of the same person are unique [9].

- Iris-based technologies have demonstrated high levels of performance, as iris is stable [8]. Moreover, it is impossible to surgically modify the pigmentation and/or colour of the iris without unacceptable risk to damage the vision.

- The physiological reaction of the iris to light sources provides one of the easiest liveness detection practices against spoofing attack.

- Iris recognition efficacy is rarely hindered by glasses or contact lenses [10]. In addition, the non-contact acquisition procedure used in capturing iris images makes it more convenient than fingerprints which mostly use optical touch based sensors.

- Among biometrics, iris has one of the smallest outlier populations, where few people cannot use or enrol using this technology [109].

Despite the aforementioned advantages of using iris recognition, the acquisition of satisfactory quality iris images for iris recognition is a critical yet challenging step [34]. It may act very poorly when deployed in the real-time applications, especially for recognition at a distance. Besides, the iris is usually located at the back of a curved and reflecting surface and typically covered by eyelashes and it is partially occluded by eyelids.

4.3 Iris Recognition System

Since the beginning of the iris recognition research, many different iris recognition systems have been developed [54, 98]. Perhaps the most successful and most well-known iris recognition algorithm, on which the state-of-the-art systems are based, is the algorithms developed by Professor John Daugman. The main stages of any typical iris recognition system include iris preprocessing, feature extraction and classification. Figure 4.2 illustrates the key phases of an iris recognition system based on the approach of Daugman [147].

The initial stage involves iris localization, iris normalization and image enhancement. The first step consists in localizing the iris area between the inner (pupillary) and outer (limbic) boundaries, with prior assumption that each border is either circular or elliptical. This process also obliges detection and removing any specular reflection, eyelash or eyelids noise from the image prior to segmentation. So as to overcome the differences in the pupil size and in the acquired images and to ensure consistency between eye images, the original segmented iris region is usually mapped into a fixed length and dimensionless pseudo-polar coordinate system. This technique is referred to as “Daugman’s Rubber Sheet” [32]. The next step is to extract distinctive features from the iris texture pattern, with the intention that comparisons between templates can be made.

Regarding feature extraction, the existing iris recognition algorithms can be classified into three major categories: phase-based image matching [34], zero-crossing representation [16] and texture analysis based approaches [185]. On the final stage, a comparison between the captured iris and the stored templates is made using matching metric. The matching metric will yield a measure of resemblance to compare the stored iris template with the claimed iris

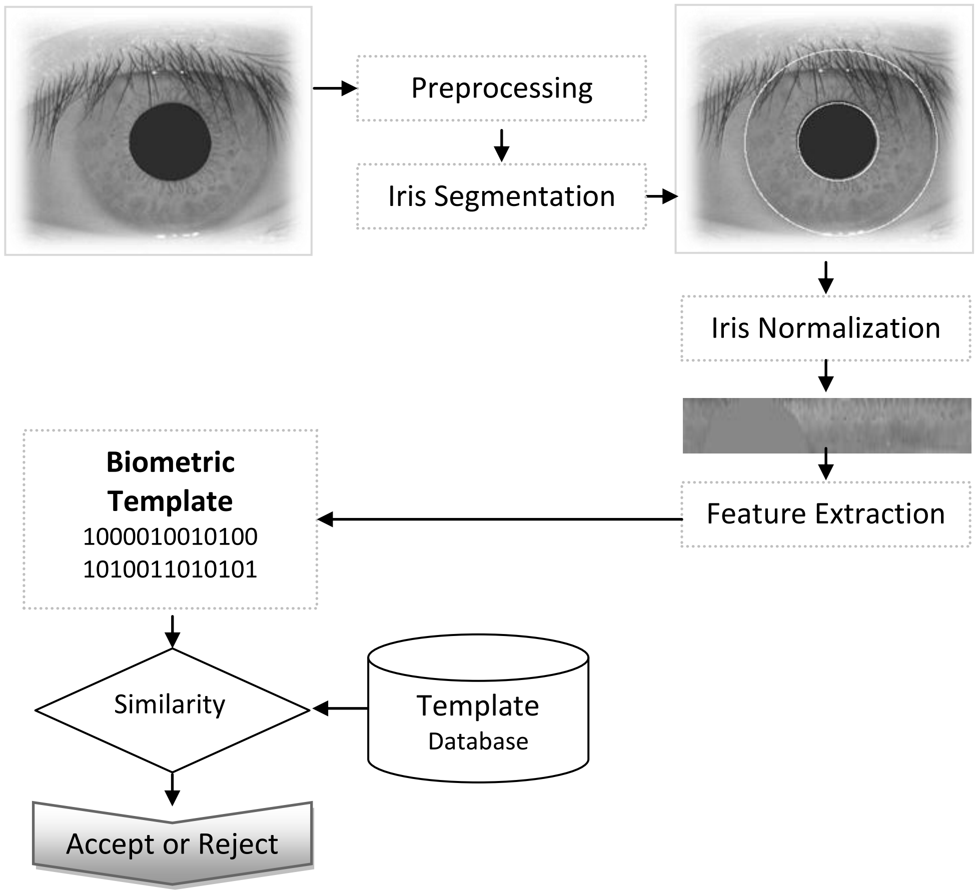


Figure 4.2 Block diagram for an iris recognition system

Quite a lot of researchers have contributed to the maturation of iris biometric technology, we will briefly review some of the key publications in this area. Daugman [32,34] applied Gabor wavelets filtering to encode the iris regions and extract the phase information of iris textures to create a 2048 bit (256 bytes) of iris template. The Hamming distance is used to compare the stored iris template with the claimed iris. Wildes et al. [185] represented another iris recognition system that decomposed the distinctive spatial characteristics of the iris into four levels Laplacian pyramid and used a normalized correlation for matching. Boles and Boashash [16] detected zero crossings of one-dimensional dyadic wavelet transform with various resolution levels over concentric circles on the iris. Both the position and magnitude information of zero-crossing representations were used to measure the similarity between the recognition and enrolment images.

Ma et al. [114] proposed an iris texture analysis method based on using multi-channel Gabor filtering to capture both global and local details in the iris. Ma et al. considered the characteristics of the iris as a sort of transient signals and identified the local sharp variation points as iris features. Lim et al. [111] used 2D Haar wavelet transform to decompose the iris image into four levels and quantized the fourth-level high-frequency information to form an 87-bit code. The researchers improved the efficiency and accuracy of the proposed system by using a modified competitive learning neural network (LVQ). Sun and Tan [176] proposal is based on using ordinal measures for iris feature representation with the objective of characterizing qualitative relationships between iris regions rather than precise measurements of iris image structures. They demonstrated that ordinal measures are intrinsic features of iris patterns and largely invariant to illumination changes.

4.3.1 Proposed Approach

As stated before, segmentation plays a crucial role in the overall achievement of the iris recognition system. Figure 4.3 below show a block diagram for the suggested approach.

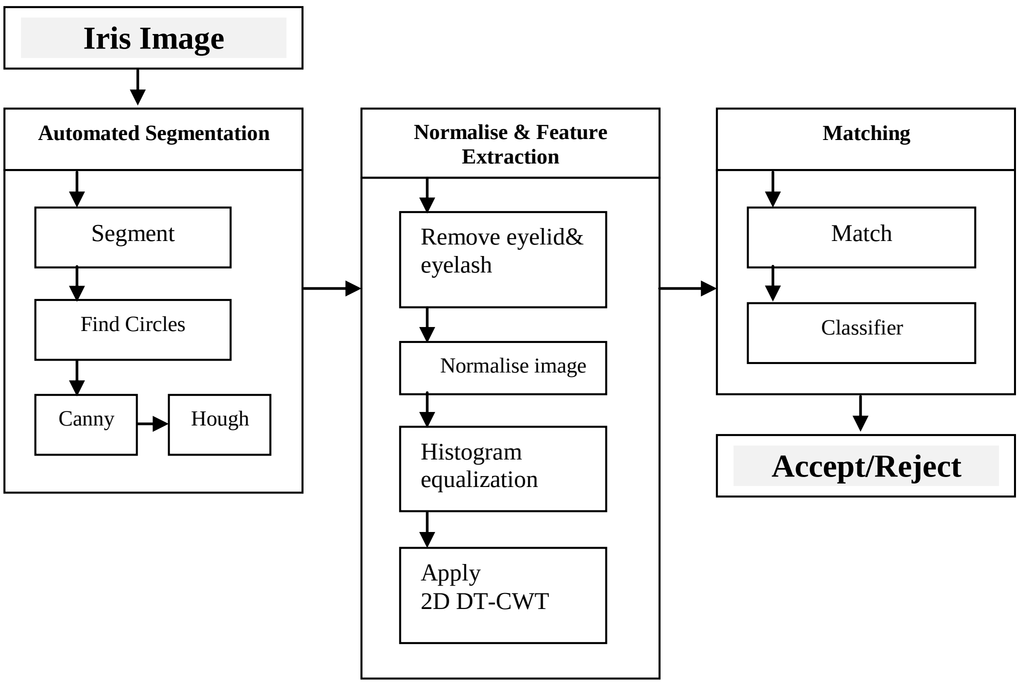


Figure 4.3 Block diagram for the suggested iris recognition approach

In the following sub-sections we will describe the proposed technique which starts with the detection of pupil and iris boundaries regions and isolating eyelids and eyelashes. Followed by extracting the features and conclude with classifying the processed iris pattern.

4.3.2 Iris Database

All the experiments in this thesis were conducted on the Chinese Academy of Sciences—Institute of Automation (CASIA) eye image database version 1.0 [21]. The CASIA iris database includes 756 frontal "non-ideal" iris images that are taken from 108 volunteers with 7 images from each person. The eye images are mainly from persons of Asian descent. The eyes of the Asian decent are characterized by their heavily pigmented irises along with dark eyelashes.

The database was collected over two sessions over a period of two months, where three samples were collected in the first session and the other four in the second session. The images were captured specially for iris recognition research using specialized digital optics. The iris images are greyscale bit-map with a resolution of 320x280. Figure 4.4 below show a number of sample images from the CASIA iris database.

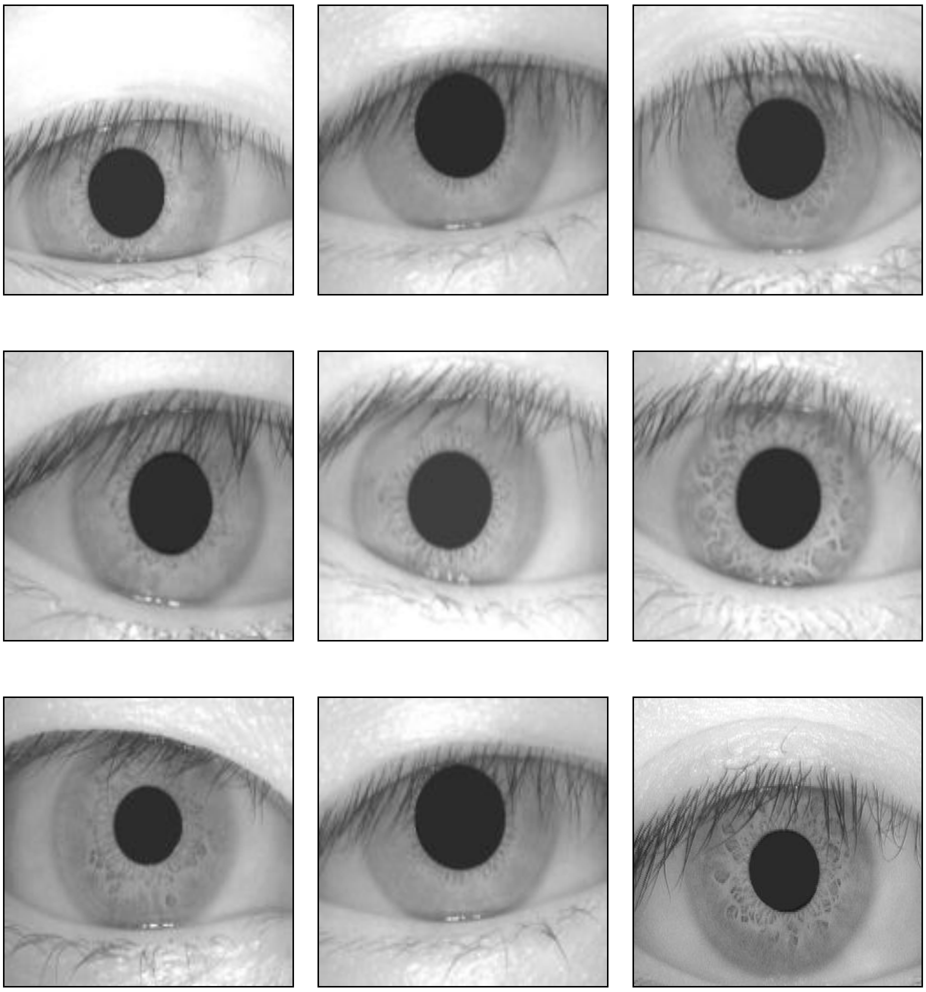


Figure 4.4 Iris image samples from CASIA V1.0 database

4.3.3 Iris Preprocessing

The primarily step in iris segmentation is to distinguish the iris texture from the input eye image. The first step in any iris recognition system is to localize the iris area between the inner (pupillary) and outer (limbic) boundaries, usually with prior assumption that each border either circular or elliptical. Researchers have proposed different algorithms for iris detection [19,31,33,34,185]. This process also obliges detection and removing any specular reflections of illumination, eyelash or eyelids occlusions from the image prior to segmentation. Segmentation plays an essential role in the overall success of any iris recognition process, as image parts that are incorrectly considered as iris pattern data will eventually lead to poor recognition rates.

4.3.4 Iris and Pupil Localization

The primary step in any iris recognition system is to localize the iris area between the inner and outer boundaries. Key steps involved involve [19]

(i) Pupil localization
(ii) Outer iris localization
(iii) Eyelids detection
(iv) Eyelashes detection.

Well-known methods such as the Integro-differential, Hough transform and discrete circular active contour models have been successfully applied in iris recognition. In the following, these methods are briefly described.

1. Daugman's Integro-differential Operator

This is by far the most cited technique and most important work [117] in the iris recognition literature. The Daugman system is patented [33] and the rights are now licensed to Iridian Technologies. The author assumes both pupil and iris has circular boundaries and applies Gaussian filter for smoothing and integration operator along the iris circle. This method tries to find a circle in the eye image with maximum change in grey level difference with its neighbours. First, due to significant contrast between iris and pupil regions the pupil boundary is localized. Then, using same operator with difference radius and parameters the outer boundary is detected. The integro-differential operator equation for detecting the iris boundary by searching the parameter space is

$$\max_{r,x_o,y_o} \left| G_\sigma * \frac{\partial}{\partial r} \oint_{r,x_o,y_o} \frac{I(x,y)}{2\pi r} ds \right| \qquad 4.1$$

where $I(x,y)$ represents the original grayscale eye image. Parameters (r, x_o, y_o) represents a circle of radius r and centre coordinates (x_o,y_o), respectively. The symbol * denotes convolution and $G_\sigma(r)$ is a radial smoothing Gaussian function with center *r* and standard deviation σ and defined as

$$G_\sigma(r) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(r-r_o)^2}{2\sigma^2}} \qquad 4.2$$

The above algorithm is applied twice, to get the boundaries of iris first then the boundaries of pupil.

2. Hough Transform

The Hough transform is a standard computer vision algorithm concerned with the identification of positions of arbitrary shapes. The conventional Hough transform was concerned with the identification of straight lines in edge-enhanced images, but later the Hough transform has been modified to identify positions of circles, ellipses and arbitrary shapes [167]. The main advantage of the Hough transform technique is that it is robust with respect to gaps in the shape boundary.

The circular Hough transform has been employed to determine the radius and centre coordinates of the pupil and iris regions by Wildes et al. [31], Kong and Zhang [95], Tisse et al. [179], and Ma et al. [115]. Wildes technique start with converting the image intensity information is into a binary edge map followed by use of a circular Hough transform [193] to localize iris boundaries. In a circular Hough transform, images are analysed to estimate the three parameters of (x_o , y_o, r) using following equations:

$$H(x_o, y_o, r) = \sum_i h(x_i, y_i, x_o, y_o, r) \qquad 4.3$$

where (x_i , y_i) is an edge pixel and i is the index of the edge pixel

$$h(x_i, y_i, x_o, y_o, r) = \begin{cases} 1 & if \quad g(x_i, y_i, x_o, y_o, r) = 0 \\ 0 & otherwise \end{cases} \qquad 4.4$$

The limbus and pupil are both modelled as circles and the parametric function *g* is defined as

$$g(x_i, y_i, x_o, y_o, r) = (x_i - x_o)^2 + (y_i - y_o)^2 - r^2 \qquad 4.5$$

The location (x_0 , y_0 , r) with the maximum value of H(x_o , y_o , r) is chosen as the parameter vector for the strongest circular boundary. Wilde’s system models the eyelids as parabolic arcs. The upper and lower eyelids are detected by using a Hough transform based approach similar to that described above. The only difference is that it votes for parabolic arcs instead of circles.

3. Discrete Circular Active Contours

Ritter proposed an active contours model to locate the pupil and iris boundaries within images [153]. First, the variance image was computed from the original image in order to improve accuracy. Afterward, an active contour model with a starting point in the centre of the pupil is initiated and moved within the iris image under the influence of using internal and external forces. The movement of the contour is based on the composition of the internal and external forces over the contour vertices. Along the active contour, the vertex moves from time *t* to time *t +1* according to

$$v_i(t+1) = v_i(t) + F_i(t) + G_i(t) \tag{4.6}$$

where v_i represents the position of the vertex at a specific time t, F_i and G_i and represent the internal and external forces, respectively. The internal forces are calibrated so that the contour forms a globally expanding discrete circle. The external forces are usually found using the edge information [117].

4.3.5 Detecting pupil and iris boundaries

Since pupil is the largest black area in the intensity image, its edges can be easily detected from the binarized image with using suitable threshold on the intensity image. With the assumption that the pupil and iris have circular shapes, Hough transformation can be used to detect edges and links edge forming iris areas especially if the shape of the object is known in advance. This involves first employing Canny edge detection technique to create an edge map. The Canny technique finds edges by looking for local maxima of the gradient of I. The gradient is calculated using the derivative of a Gaussian filter. This technique uses two thresholds, to detect strong and weak edges, and includes the weak edges in the output only if they are connected to strong edges. This technique is therefore less likely than the others to be fooled by noise, and more likely to detect true weak edges [118].

As suggested by Wildes et al. [185] Gradients were biased in the vertical direction for the outer sclera boundary. Whereas, the vertical and horizontal gradients were weighted equally for the inner pupil boundary. The reason behind using the vertical coefficients when performing a circular Hough transform for detecting the outer sclera boundary is that it should reduce the influence of the eyelids since eyelids are usually horizontally aligned.

To increase the efficiency and accuracy of the circle detection process, the Hough transform was performed first for the sclera boundary, then for the pupil boundary within the iris region. Figure 4.5 contains some of the images of localized irises from the CASIA database.

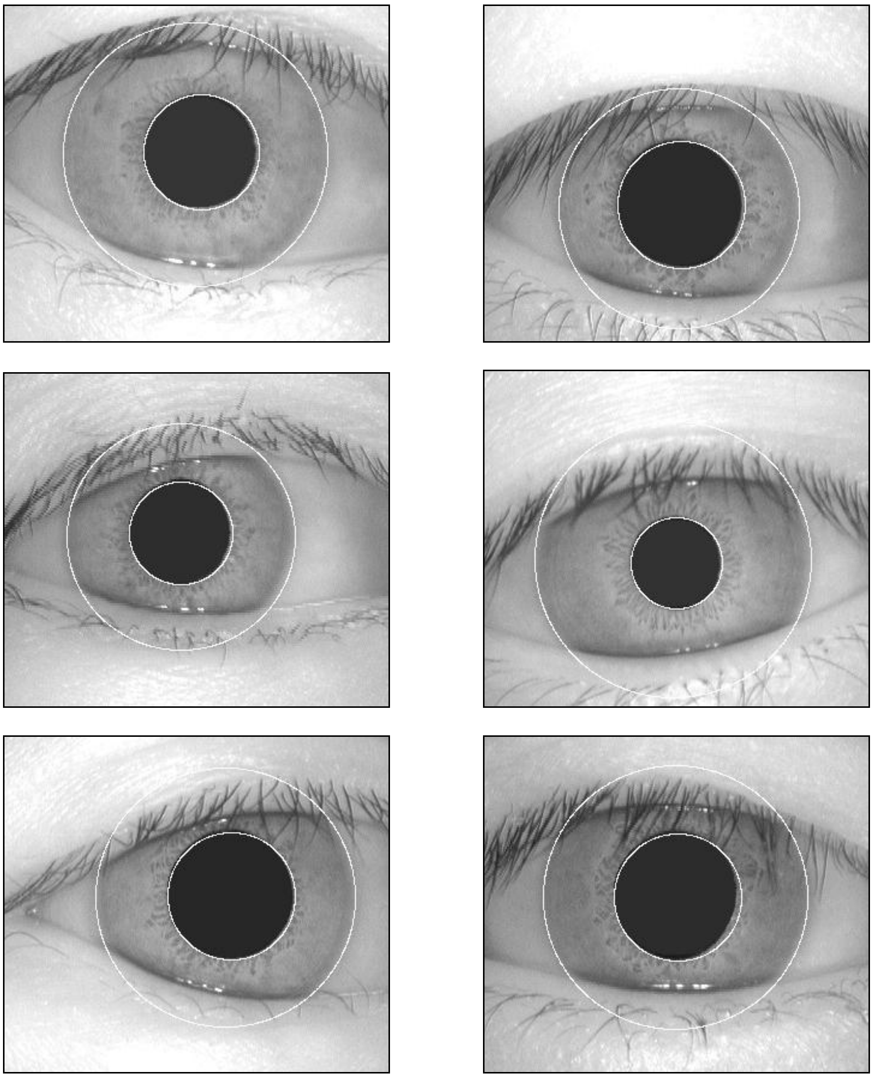


Figure 4.5 Illustration the results of the proposed iris segmentation technique

4.3.6 Isolating eyelids and eyelashes

Processing iris images is a challenging task since the iris region can be occluded by eyelids or eyelashes. This will cause a significant difference between the intra- and inter-class comparisons. Conventional techniques for isolating eyelids and eyelashes have several drawbacks. Firstly, the process of detecting the eyelids and eyelashes is complex and computationally expensive. Secondly, conventional techniques demand extra memory requirements to store the generated noise mask for each iris template.

Lastly, the classification accuracy is expected to be degraded as several tracks that existed nearby the eyelid or pupil regions are badly corrupted. Therefore, we decided to isolate the effect of the eyelids and eyelashes by using only the left and right parts of the iris area for the iris recognition. Most of the methods extract the complete iris image, but we plan to exclude these parts of the iris image for recognition. The process of detecting the eyelids and eyelashes is depicted in Figure 4.6 below.

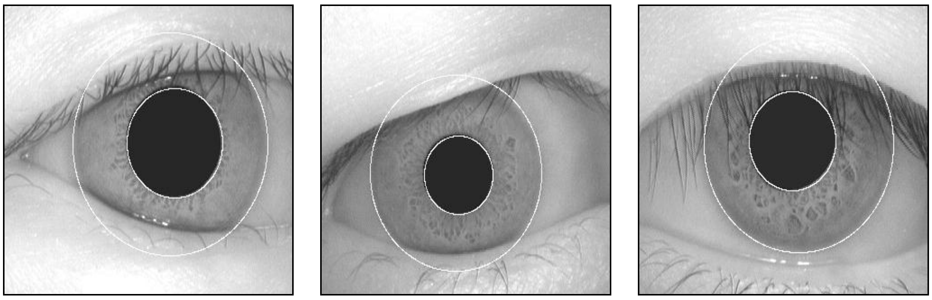
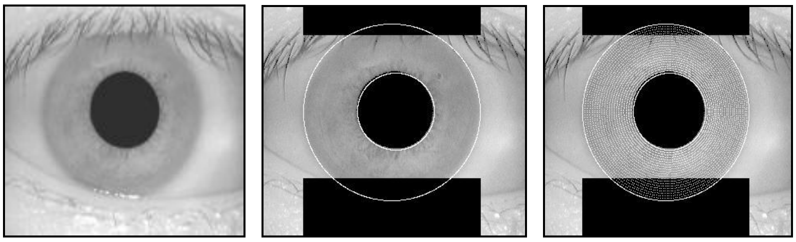


Figure 4.6 Figure Examples of extracted iris area occluded by the eyelashes and/or upper and lower eyelids

Eyelids were first detected by first fitting a line to the upper eyelid using the linear Hough transform. A second horizontal line is then drawn, which intersects with the first line at the iris edge that is closest to the pupil [117]. This process is illustrated in Figure 4.6 and is done for both the top and bottom eyelids. The second horizontal line allows highest isolation of eyelid regions. Canny edge detection is used to create an edge map, and only horizontal gradient information is taken. If the maximum in Hough space is lower than a set threshold, then no line is fitted, since this corresponds to non-occluding eyelids. Besides, the lines are constrained to lie outside the pupil region, and inside the iris region.

The process is concluded by trimming the iris area above the upper boundary of the pupil and the area below the lower boundary of the pupil. Figure 4.7 illustrates the proposed technique.



(a) (b) (c)

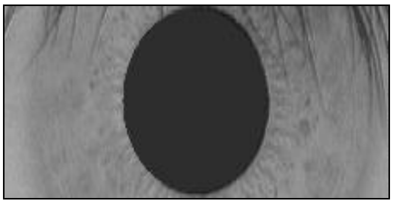
Figure 4.7 Example of localized iris where the upper and lower parts is occluded and the segmentation result, black regions denote detected eyelids and eyelashes regions. (a) original image, (b) and (c) localized iris region

Afterward, we apply histogram equalization to enhance the contrast of segmented iris images [118]. Histogram equalization method is widely used in image processing, in order to enhance the images global contrast of images by adjusting image intensities. Through this adjustment, it reassigns the intensity value of the pixels based on the image histogram. This process assigns the intensity values of the input image such that the output image contains a uniform distribution of intensities.

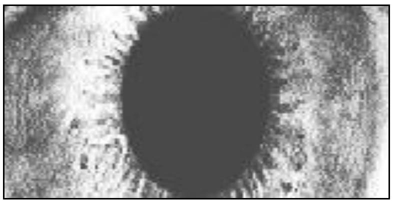
Let j be the intensity value of a pixel in the original image. The new value k is given by:

$$k = \sum_{i=0}^{j} \frac{N_i}{T} \qquad 4.7$$

where N_i is the number of pixels with intensity value i and T is the total number of pixels that the image contains. Figure 4.8 shows the result of histogram equalization.



(a)



(b)

Figure 4.8 Result of histogram equalization

(a) localized iris region (b) localized iris image after histogram equalization

Once the iris region is segmented, the next step is to eliminate the translation variance by moving the centroid of the image to the centre of the iris image. For that reason, the image is normalized so that it fits into a resolution of 156x100 pixels. Example of the result of this process is shown in Figure 4.7.

4.4 Feature extraction

The iris has fascinating texture information. Therefore, it is attractive to search representation methods which can capture the local crucial information in an iris. There have been many techniques suggested in the literature for extracting unique and invariant features from the iris image. These techniques can employ either texture- or appearance-based features. An in-depth comparison of these two approaches, as well as information on several other less-well-known approaches, can be found in [19].

Wavelet techniques are successfully applied to a wide range of problems in signal processing, classification, data compression and denoising. Researchers in the iris recognition field have used a range of wavelets to analyse the iris texture [160,176, 178]. The wavelet transform is a very powerful tool for structural texture analysis [137]. It is a linear operation that decomposes a signal into components that appear at different scales. Such decomposition has been thoroughly studied in signal processing and computer vision. For a more comprehensive description, the reader is referred to [1,160].

Wavelet transform is based on the convolution of the signal with a dilated filter. However, it is well known that the ordinary discrete wavelet transform is not shift-invariant because of the decimation operation during the transform. Therefore, any minor shift in the input signal can cause very different output wavelet coefficients. Moreover, ordinary discrete wavelet transform (DWT) is not appropriate for the analysis of high-frequency signals with relatively narrow bandwidth. To overcome some of the shortcomings of the DWT, Kingsbury [87] introduced the dual-tree complex wavelet transform (DT-CWT).

2D Dual-Tree Complex-Valued Wavelet for Iris Analysis

DT-CWT has improved directionality and reduced shift sensitivity and it is approximately orientation invariant [163]. The DT-CWT consists of real parallel wavelet transforms pair where the wavelets of one branch are the Hilbert transforms of the wavelets in the other. In this case, the wavelets in the two trees of the DT-CWT can be considered as the real and imaginary parts of complex coefficients. Accordingly, any input image can be decomposed into its 6 directional subbands. At each scale, the DT-CWT generates 6 directional subbands with complex coefficients, oriented at ±15°, ±45°, and ±75°. The real (R_i) and imaginary (C_i) parts of an impulse responses of the complex wavelets filters under 6 directional subbands are illustrated in Figure 4.9.

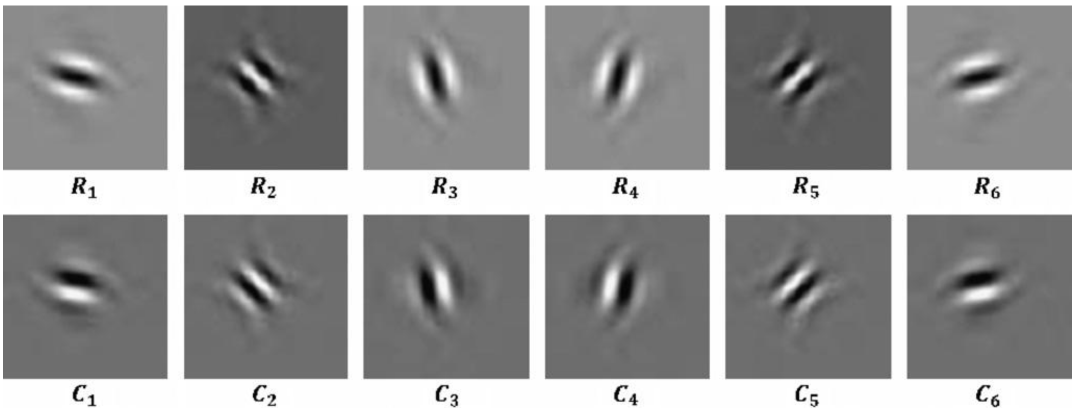


Figure 4.9 Complex dual-tree 2D wavelets and corresponding labels, *adapted from* [163]

The one-dimensional DT-CWT decomposes the input signal *f(x)* by expressing it in terms of a complex shifted and dilated mother wavelet ψ(x) with associated scaling function *ϕ(x)*, is defined as,

$$f(x) = \sum_{l \in Z} S_{j_o,l}\phi_{j_o,l}(x) + \sum_{j \geq j_o}\sum_{l \in Z} c_{j,l}\psi_{j,l}(x) \qquad 4.8$$

where Z is the set of natural numbers, *j* and *l* refer to the index of scale and transition factor respectively, $s_{j0,l}$ is the scaling coefficient and $c_{j,l}$ is the complex wavelet coefficient with

$\phi_{j_o,l}(x) = \phi^{r}{}_{j_o,l}(x) + \sqrt{-1}\phi^{i}{}_{j_o,l}(x)$ and $\psi_{j,l}(x) = \psi^{r}{}_{j,l}(x) + \sqrt{-1}\psi^{i}{}_{j,l}(x)$, where the superscripts r denote the real part and i the imaginary part. The set $\left\{\emptyset^{r}_{j_{0,l}}(x), \emptyset^{i}_{j_{0,l}}(x), \psi^{r}_{j_{0,l}}(x), \psi^{i}_{j_{0,l}}(x)\right\}$, in the 1D DT-CWT forms a tight wavelet frame with a redundancy factor of two.

The final transformed real and imaginary coefficients of the 1D DT-CWT are computed using separate filter banks on parallel working on the same data with filters h_0 and h_1 for the real part, and g_0 and g_1 for the imaginary part, as illustrated in Figure 4.10 [163].

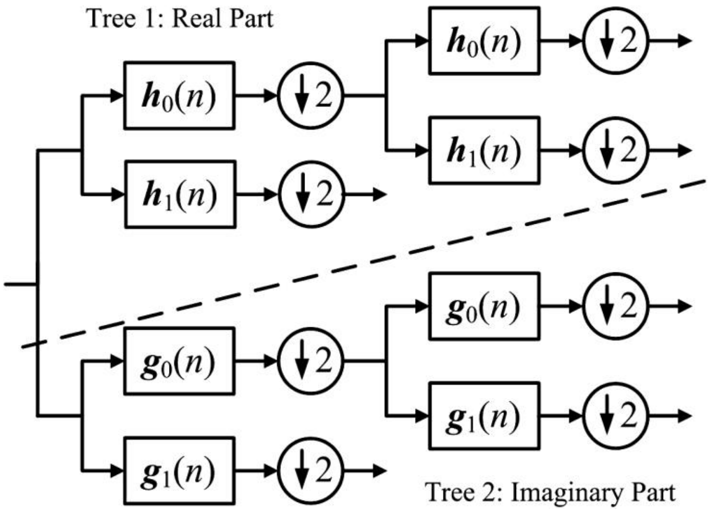


Figure 4.10 One-dimensional DT-CWT filterbank implementation to obtain real parts: h_0 and h_1 and imaginary parts: g_0 and g_1 for 1D signal, *adapted from* [87]

The two-dimensional DT-CWT decomposes a 2D image *f(x,y)* through a series of dilations and translations of a complex scaling function and six complex wavelet functions $\psi^{\theta}_{j,l}(x)$, oriented in angles of $\theta = \{\pm 15°, \pm 45°, \pm 75°\}$, i.e.,

$$f(x,y) = \sum_{l \in z^2} s_{j_0,l} \emptyset_{j_0,l}(x,y) + \sum_{\theta \in} \sum_{j \geq j_o} \times \sum_{l \in z^2} c^{\theta}_{j,l}\, \psi^{\theta}_{j,l}(x,y) \tag{4.9}$$

Thus, the decomposition of *f(x,y)* by exploiting the DT-CWT gives with one complex-valued low-pass subband and six complex-valued high-pass subbands at each level of decomposition, where each high-pass subband at the angles of $\{15^o, 45^o, 75^o, 105^o, 135^o, 165^o\}$.

4.5 Classification Stage

In the field of iris recognition researchers have used a variety of wavelets to evaluate the iris texture [19]. Some techniques used the output of the wavelet transform to create a binary feature vector, by quantizing each real value into binary form by converting the positive value into 1 and the negative value into 0. Though, others have considered using the real-value output in building the feature vector, in our approach we have chosen to obtain the real-value output of decomposition level and use it to feed the classifier model.

Regarding the classification algorithm, the most important thing is its capacity to discriminate, based on the available information. SVM has been chosen since it proven advantageous in handling large scale classification tasks with good generalization performance. Additionally it has demonstrated superior results in various classification and pattern recognition problems [63]. Furthermore, for several pattern classification applications, SVM has already been proven to provide better generalization performance than conventional techniques especially when the number of training samples is small and the number of input variables is large.

With this purpose in mind, we evaluated the SVM against two unsupervised classification algorithms: k-NN Naïve Bayes. In this section we will offer brief background knowledge on SVM.

4.5.1 Overview of SVM

SVM has been recently proposed as a popular tool for solving many classification tasks based on the statistical learning theory invented by Vapnik [181]. For this purpose we turn to SVM for validating our approach. SVM is the interest in this study for its good classification accuracy reported in many pattern recognition problems. To achieve better generalization performance of the SVM, original input space is mapped into a high-dimensional dot product space called the *feature space*, and in the feature space the optimal hyperplane is determined. The optimal hyperplane is found by exploiting the optimization theory, and respecting insights provided by the statistical learning theory.

Linearly separable data

Given training vectors x_i, i=1, … , N of length n and a vector y defined as follows

$$y_i = \begin{cases} 1 & if \quad x_i \quad in \quad class \quad 1 \\ -1 & if \quad x_i \quad in \quad class \quad 2 \end{cases} \qquad 4.10$$

The central idea of SVM is to define a separating hyperplane, so that the classification margin between the two classes is as large as possible, measured along a line perpendicular to the hyperplane.

The SVM training paradigm finds the separating hyperplane which gives the maximum margin or distance between the parallel hyperplanes that are as far apart as possible while still separating the data. These hyperplanes should satisfy the following constraints. Since the wider margin can acquire the better generalization ability.

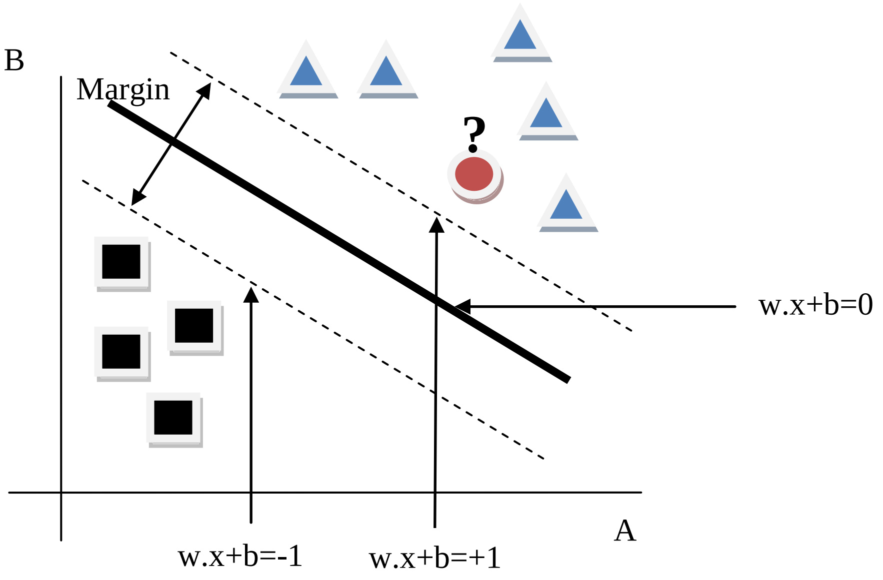


Figure 4.11 Linearly separable data
▲: Class 1, y=+1 ■: Class 2, y=-1

In Figure 4.11, two class's instances could be separated by bold solid line. The test sample (the circle) can be classified based on the hyperplane. In this figure, the hyperplane that is calculated from these training examples is given by the bold line, separated from the closest training vectors by the distance d. The classification of an unknown sample is done by determining which side of the hyperplane the new instance falls. In this example, the prediction for the unknown sample would be triangle.

So we can define a canonical hyperplanes as follows (Vapnik, 1995):

$$\begin{cases} H_1 = w^T x + b = +1 \\ H_2 = w^T x + b = -1 \end{cases} \qquad 4.11$$

In addition, all training samples x_i satisfy:

$$\begin{cases} w^T x + b \geq 1 & for \quad y_i = +1 \\ w^T x + b \leq -1 & for \quad y_i = -1 \end{cases} \qquad 4.12$$

For linearly separable data, any hyperplane *g(x)=0* can be written as

$$g(x_i) = w^T x + b = 0 \qquad 4.13$$

where *w* is an n-dimensional vector , *b* is the offset of the hyperplane from the origin and *x* represents n-dimensional vector representing any point on the hyperpalne. The vector *w* and the scalar *b* determine the position of the separating hyperplane. The distance between each of the canonical hyperplanes and the separating hyperplane is $\frac{1}{|w|}$. Now maximizing the separating margin is equivalent to maximizing the distance between hyper plane H_1 and H_2. Hence we can get the maximal width between them $m = (x^+ - x^-).\frac{w}{\|w\|} = \frac{2}{\|w\|}$. Now we can formulate the learning problem of SVM to maximize the margin the task as follows

$$\min imize \qquad g(w) = \frac{1}{2}\|w\|^2 \qquad 4.14$$

So that: $w^T x_i + b \geq 1, \quad \forall i$

This enable us to use the Lagrange formalism to obtain the *primal form* of the objective function L_p , which is

$$\min imize \qquad L_p(w, b, \alpha_i) = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{n} \alpha_i (y_i (w^T x_i + b) - 1 \qquad 4.15$$

where $\{\alpha_i : 1, ..., n; \alpha_i \geq 0\}$ are the Lagrange multipliers.

Solving the minimization problem is equivalent to finding the values *w*, *b*, and $\alpha_i \geq 0$ that minimize L_p. To do so, we initial differentiate L_p with respect to *w* and *b*. Then, by equating the derivates to zero we get

$$\frac{\partial L_p}{\partial b} = 0 \Rightarrow \sum_{i=1}^{n} \alpha_i y_i = 0 \qquad 4.16$$

$$\frac{\partial L_p}{\partial w} = 0 \Rightarrow w = \sum_{i=1}^{n} \alpha_i y_i x_i \qquad 4.17$$

when differentiating with respect to *b* and *w* respectively.

Taking these two equalities and substituting into L_p yields the dual form of the Lagrangian. We want to maximize

$$L_p = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\ \sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j x_i^T x_j \qquad 4.18$$

subject to

$$\sum_{i=1}^{n}\alpha_i y_i = 0, \qquad \alpha_i \geq 0 \qquad 4.19$$

This optimization formulation is expressed using inner product of the training samples x_i and the numbers of training samples n.

Linearly non-separable data

In the previous section the SVM theory was introduced as an optimization problem, under the assumption that the data are linearly separable. However, in many practical problems, data is subject to noise or outliers, so it is impossible to draw linear boundaries between classes. Hence, in order to extend the support vector theory to solve imperfect separation, positive slack variables is introduced $\{\xi_i : i,1,...,n; \xi_i \geq 0\}$ into the original constraints [181] along with an additional penalty value C for the points that cross the boundaries to consider the misclassification errors. C is a regularization parameter used to decide a trade-off between the training error and the margin. If C is chosen too small, it may cause the problem of under-fitting of the training data. If C is too large, the algorithm may increase the possibility of over-fitting.

$$\min imize \qquad g(w,\xi) = \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \qquad 4.20$$

so that

$$y_i(w^T x_i + b) \geq 1 - \xi_i, \xi_i \geq 0 \qquad 4.21$$

The primal and dual forms of the Lagrangian are built as

$$L_d = \sum_{i=1}^{n}\alpha_i - \frac{1}{2}\sum_{i=1}^{n}\ \sum_{j=1}^{n}\alpha_i\alpha_j y_i y_j x^T{}_i x_j \qquad 4.22$$

subject to

$$\sum_{i=1}^{n}\alpha_i y_i = 0, \qquad 0 \leq \alpha_i \leq C \qquad 4.23$$

Kernel-trick

The initial optimal hyperplane algorithm proposed by Vapnik [181] was a linear classifier. Yet, Boser et.al [18] suggested a way to create nonlinear classifiers by applying the kernel trick to to extend the linear learning machine to handle nonlinear cases. Kernel function is essentially a weighted function designed for nonparametric function estimations. We aimed to maximize the margin of separation between patterns to have a better classification result. The calculations can be simplified by converting the problem with Kuhn-Tucker conditions into equivalent Lagrange dual problem.

With this mapping, the discriminant function is of the follow form

$$g(x_i) = w^T \phi(x) + b \tag{4.24}$$

And the dual form of the Lagrangian becomes

$$L_d = \sum_{i=1}^{n} \alpha_i - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \alpha_i \alpha_j y_i y_j \phi(x_i)^T \phi(x_j) \tag{4.25}$$

subject to

$$\sum_{i=1}^{n} \alpha_i y_i = 0, \qquad C \geq \alpha_i \geq 0 \tag{4.26}$$

Overall, any positive semi-definite functions $K(x_r.x_i)$ that satisfy Mercer's condition can be kernel functions. The function $K(x_r.x_i)$ that returns a dot product of two mapped patterns is called a kernel function.

Different kernels can be selected to construct the SVM. The most commonly used kernel functions are the polynomial, linear and Gaussian radial basis kernel function (RBF).

- **Linear kernel function:**

$$k(x_i, x_j) = (x_i^T x_j) \tag{4.27}$$

- **Gaussian RBF:**

$$k(x_i, x_j) = \exp(-\gamma \left\| x_i^T x_j \right\|^2), \gamma > 0 \tag{4.28}$$

- **Polynomial kernel function:**

$$k(x_i, x_j) = (r + \gamma x_i^T x_j)^d, \gamma > 0 \tag{4.29}$$

where γ , *r* and *d* are kernel parameters.

4.6 Recognition Results

In this section, experiment is performed in order to evaluate the performance of the proposed scheme.

Since the dual-tree complex wavelet has the properties of shift invariance and multi-resolution representation, we perform the 2D dual-tree complex wavelet on the normalized images and combined the features at different resolution scales to form a feature vector to train and test the classifiers. In general, dual-tree complex wavelet contains both real and imaginary terms [87]. However, in our research, in order to reduce processing time and complex operations, the iris feature vector consists of only the real part from the highest level as shown in Equation 4.9. So as to alleviate the demand of large computational burden and high memory requirement of the dual-tree complex wavelet-based iris recognition and at the same time retain most of its desired properties, the directional multi-scales decomposition of the normalized iris image are performed up to the 6^{th} level of decomposition as described in the below table.

Table 4.1 The dimension of feature vector after applying various 2D DT-CWT decomposition levels

	decomposition level			
	3^{rd} Level	**4^{th} Level**	**5^{th} Level**	**6^{th} Level**
Dimension	1040	280	80	24

The classification experiments involve two main steps. Firstly, the classifiers need to be trained with labelled samples in order to be able to perform verification. Secondly, the trained classifiers need to be tested with unlabelled samples to determine their classification accuracy using 10-fold cross-validation. The implementation is carried out via LIBSVM tool version 2.6, which is initially designed by Chang and Lin [23]. LIBSVM is an integrated software package for support vector classification, regression, and distribution estimation. It supports multiclass classification. The basic algorithm uses the sequential minimal optimization (SMO) for the multi-class SVM.

4.6.1 Parameter Selection of SVM

The effectiveness of SVM depends on kernel used, kernel parameters and a proper soft margin or penalty C value [69]. The selection of a kernel function is an important problem in applications although there is no theory to tell which kernel to use. Selection of the kernel, perhaps from among the presented kernels, is usually based on experience and knowledge about the classification problem at hand.

Gaussian RBF kernels have been found to be the most powerful amongst the above mentioned kernels [38]. Moreover, the RBF requires less parameter to set than a polynomial kernel. However, convergence for RBF kernels takes longer than for the other kernels [133]. Overall, RBF and other kernel functions have similar overall performance.

In developing techniques for efficient parameter selection, Hsu et al. have proposed a procedure to get acceptable yet reasonable results with LIBSVM [69]. To get appropriate generalization ability, we conduct a validation process to choose parameters. The procedure is as the following [69]

1. Consider a grid space of (C,γ) with log2 $C \in$ {-10,-9, . , 4} and log2 $\gamma \in$ {-2,-1, . . . , 12}.
2. For each hyper-parameter pair (C,γ) in the search space, the validation performance is measured by conducting 5-fold cross validation on the training set.

3. Choose the optimal parameters pair (C, γ) that leads to the highest cross-validation accuracy.
4. Use the best parameters to build the SVM model.

The discriminating features are extracted from the transformed image using 2D DT-CWT at different resolution scales and the extracted features are used to train the SVM. The kernels used in our experiments include the Gaussian RBF kernel, the Polynomial kernel and the Linear kernel. Table 4.2 and Figure 4.12 summarize the classification rates using three SVM kernel functions with different decomposition scales.

Results indicate that the Gaussian RBF kernel function performed equally well or suppressed the performance of the other kernel functions in recognition rate where the best accuracy rate of classification was 92.86% at the third level of decomposition.

Table 4.2 The recognition rates (%) of the proposed method by using different SVM kernels

Kernel function	**Scale**			
	Level 3	**Level 4**	**Level 5**	**Level 6**
Gaussian RBF	**92.86**	**91.79**	**85.25**	**71.95**
Polynomial	92.46	91.66	84.78	71.29
Linear	92.32	91.53	84.25	67.72

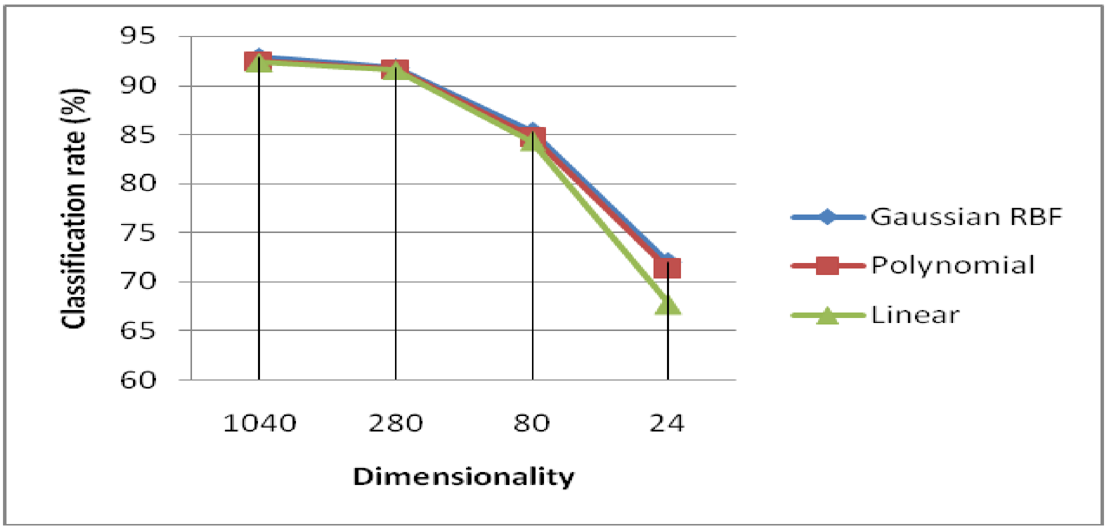


Figure 4.12 Classification rate among SVM kernels vs. dimensionality with different number of decomposition levels

The classification accuracy of the SVM is also compared with the k-NN algorithm and Naïve Bayes classifiers. The classification accuracies of SVM, k-NN and Naïve Bayes classifiers are shown in Figure 4.13 and Table 4.3.

Table 4.3 Iris recognition performance for 2DT-CWT with different number of decomposition levels

Decomposition Level	2DT-CWT Dimension	Best SVM	k-NN	Naive Bayes
3	1040	**92.86**	**80.82**	75.26
4	280	91.79	78.96	**77.11**
5	80	85.25	73.67	71.29
6	24	71.95	55.95	56.74

In all experiments, SVM outperformed the performance of the other classifiers in recognition rate when equal number of decomposition scales is used. The highest recognition rate we achieve is **92.86**% at the third level of decomposition with a feature vector of 1040.The best classification rate achieved by the k-NN classifier was **82.82**% at the third level of decomposition. Whereas the Naïve Bayes classifier achieved its best performance at the fourth level with an accuracy of **77.11**%.

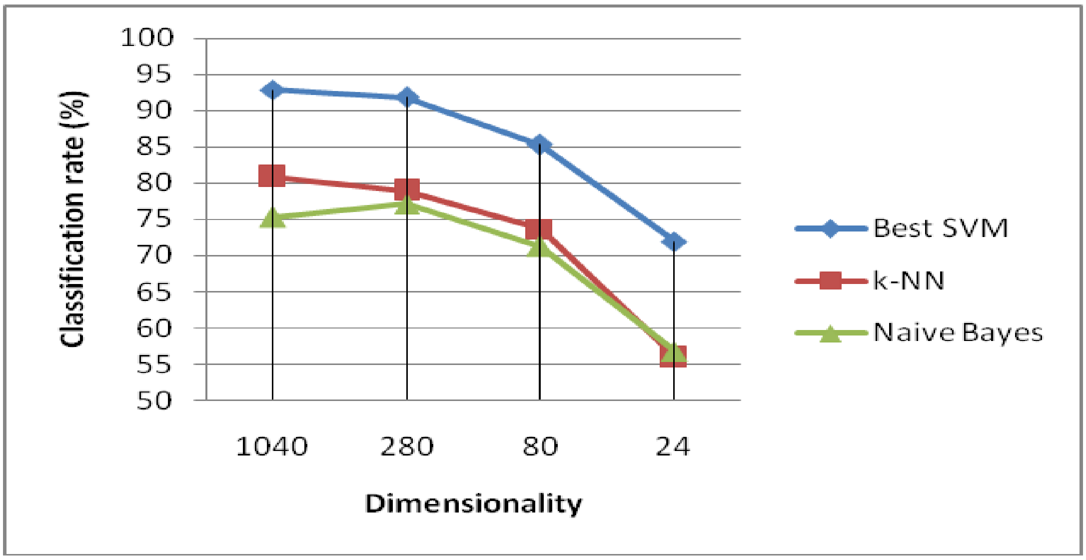


Figure 4.13 Classification rate vs. dimensionality for 2DT-CWT with different number of decomposition levels

The accomplished results indicate also that the SVM is more effective than other conventional classifiers even when the input dimension space is high. It is clear that the dual-tree complex wavelet features are very stable in iris recognition. The success of the dual-tree complex wavelet is due to its approximate shift invariant property and its good directional selectivity in 2D.

4.6.2 Comparison with Existing Methods

In the experiments, we compared proposed technique with those proposed by Wildes [185],Narote and Narote [131] , Chen and Yuan [26] and Masek [117]. They were chosen due to the fact that they have reported results using the same iris database.

Table 4.4 lists the classification rates of the other techniques against the proposed method. It is clear that proposed technique achieves good recognition rate. The results showed that the complex wavelet based representation is as discriminating as other techniques. With 1040 features, the recognition rate of 2D DT-CWT combined with the SVM is over 90%.

Table 4.4 Comparison of Recognition Performance on CASIA 1.0 Iris Database

Methodology	Accuracy rate %
Wildes [185]	86.49
Masek [117]	83.97
Chen and Yuan [26]	91.80
Narote and Narote [131]	91.33
Proposed	**92.86**

4.7 Summary

This chapter proposes new iris segmentation approach based on minimizing the effect of the eyelids and eyelashes by trimming the iris area above the upper and the area below the lower boundaries of the pupil. The 2D DT-CWT is extracted from the iris images and used to increase the recognition accuracy.

The comparison of proposed features is evaluated on the diverse classification schemes; Naïve Bayes, k-NN and SVM. Our experimental results indicate that the SVM classifier indicates that its performance is generally the best of all the classifiers evaluated in this paper. Among the used SVM kernels, the Gaussian RBF kernel function is the best for iris recognition in our experiments. Experimental results also indicate that the performance of SVM as a classifier is far better than the performance of k-NN and Naïve Bayes classifiers.

The proposed innovative technique is computationally effective as well as reliable in term of recognition rate of 92.867% compared with other techniques. The combination of dual-tree complex wavelet with SVM is promising.

Chapter 5 Feature Fusion of Online Signature and Iris Biometrics

Unimodal biometric systems that based on utilising a single biometric trait often face practical limitations that negatively influence their overall performance. This is expected to a variety of reasons such as noisy data, intra-class variability, low distinctiveness, non-universality and unacceptable error rates due to the nature of relevant biometric traits [80]. Multimodality, that is the integration of several biometric traits for accurate authentication, is often seen as a way to solve some of the aforementioned limitations [154]. The efforts in the area of biometric authentication have been directed toward the fusing the information obtained from a range of independent modalities. Multimodal approach relies on fusing separate information from different modalities to provide complementary information to achieve a more reliable recognition of individuals. For example, a common approach is to combine face and speech modalities to achieve a more trustworthy recognition decision [13,51,90,145,146].

As mentioned earlier in Chapter 2, fusion in multimodal systems can take place at four possible levels sensor, feature, matching and decision. The sensor and the feature levels are referred to as a *pre-mapping* fusion while the matching score and the decision levels are referred to as a *post-mapping* fusion [161]. In pre-mapping fusion, the data is integrated before any use of classifiers. While in post-mapping fusion; the data is integrated after mapping into matching score/decision space. The matching score-level fusion approach has attracted a lot of attention although that the amount of information available for fusion declined progressively after each layer of processing in a biometric system [41]. We have observed that, only limited work is reported on feature level fusion of multimodal biometric system [149,154].

In his Chapter we are going to suggest a number of fusion schemes at the feature level and we limit ourselves to two modalities, namely, iris and online signature. We expect that the accuracy of the combined biometrics is going to be better than unimodal systems based on iris [4] or handwriting signature [7] alone.

Therefore, we aim to answer the following questions: which fusion strategy can bring the best results in terms of performance and how much improvement can we expect from a feature fusion scheme? Toward this objective, we will design several feature fusion schemes at different possible feature levels. Moreover, we will also address the complexity problem regarding feature space, in the sense that we will also raise the question whether it could be possible to reduce the dimension of the fusion feature space, through an appropriate selection procedure, while keeping the same level of performance.

The rest of this chapter is organized in the following manner. Section 5.1 provides an overview of the current multimodal biometric research. Sections 5.2 and 5.3 describe the concept of feature fusion and the different architectures that we have designed for iris and online signature feature level fusion. Then, Section 5.4 reports the comparative results obtained using the different architectures and summarizes the main results of this chapter and finally offers concluding remarks.

5.1 Multimodal Biometrics Authentication

Several approaches have been proposed and developed for the multimodal biometric authentication system. Ben-Yacoub et al. [13] evaluated five binary classifiers on combinations of frontal face image and speech modalities (XM2VTS database). They found that SVM and bayesian classifier achieved almost the same performances and both outperformed Fisher’s linear discriminent, C4.5 decision tree and MLP. The Linear Weighted classifier has outperformed the Linear SVM, but the SVM is demonstrated to have possessed an advantage in combining potentially any number of modalities at the same computational cost with very good fusion results.

The use of hybrid biometric person authentication based on face and voice features has been explored in a study presented in [146]. Although a simple logical AND scheme is used for the purposes of fusion, the experimental results have confirmed that a multimodal approach is better than any single modality.

A commercial multimodal system called BioID based on the fusion the scores or decisions of face, voice and lip movement was proposed by Frischholz et al. [51]. Lip motion and face images were extracted from a video sequence and the voice from an audio signal. Four different score-level fusion methods and one decision-level fusion method are empirically compared in that study. However, their algorithm is restricted to only the AND and OR rules. Accordingly to the security level, experiments conducted on 150 subjects demonstrated a decrease below 1% of the FAR.

In 2003, Fierrez-Aguilar et al. developed a multimodal approach including a face verification system based on a global appearance representation scheme, a minutiae-based fingerprint verification system and an online signature verification system based on HMM modeling of temporal functions. The scores are combined by means of SVM classifiers, from which user-independent and user-dependent strategies are applied at the score level [47]. Results indicated that appropriate selection of parameters for the learning-based approach has delivered better verification performance than the rule-based approach. The EERs of the unimodals of face, online signature, and fingerprint verification systems were 10%, 4% and 3%, respectively. Results showed that the Sum Rule reduced the EER to 0.5% and the RBF SVM fusion strategy reduced the EER to 0.3% and 0.05% respectively for the user-independent and user-dependent fusion strategies.

Also, in that year, Kumar et al. [97] proposed a multimodal approach for palmprint and hand geometry images. Two schemes of fusion were applied, one at the feature level by concatenating the feature vectors, and the other at the matching score level by max rule. Only the fusion approach at the matching score level outperforms the unimodal systems. The multimodal approach obtained a FAR of 0% and a FRR of 1.41%, while the best unimodal approach in this study, the palmprint-based verification system, obtained a FAR of 4.49% at an FRR of 2.04%.

Ross and Jain, proposed a multimodal system combined the biometrics of face, fingerprint and hand geometry with three fusion techniques at the matching score level. They applied sum rule, decision trees, and linear discriminant function, after normalizing the scores [158]. The approach with the sum-rule fusion method outperforms the other fusion strategies, as well as the unimodal systems. At a FAR of 0.03%, the combination approach obtained a FRR of 1.78%.
Wang et al. proposed a multimodal approach for a PCA-based face verification system and a key position local variation-based iris verification system, with fusion methods at the two matching scores using unweighted and weighted sum rules, Fisher discriminant analysis, and neural networks with radial basis function (RBFNN) to [182].

In 2004, Toh et al. [180] fingerprint, hand geometry and voice biometrics were integrated using weighted-sum-rule based match-score-level fusion. They addressed the multimodal decision fusion problem as a two-stage problem: learning and decision. They introduced a reduced multivariate polynomial model to overcome the tedious recursive learning problem in multimodal biometrics in order to achieve good decision accuracy. Four global and local learning and decision paradigms were suggested and explored to observe their decision capability. The four learning and decision paradigms were investigated, adopting the reduced polynomial model for biometric decision fusion. Experiments showed that local learning alone can improve ERRs of about 50%. The have noticed that local decision can be improved once threshold settings are appropriately selected for each user.

Ross and Govindarajan [159] proposed a multimodal biometric system utilising Face and hand geometry at feature level. Face was represented using PCA and LDA while 32 distinct features of hand geometry is extracted and then concatenated to form a fused feature. After that, Sequential Feed Forward Selection (SFFS) was employed to select the most valuable features from the fused feature space. In 2005, Snelick et al. [171] investigated the performance of integrating three fingerprint recognition commercial systems and one face recognition commercial system multimodal biometric systems using a population of 1,000 individuals, at the score level. Seven score normalization techniques (min–max, z-score, tanh, adaptive, two quadrics, logistic, and quadric-line-quadric) and five fusion techniques on the normalized scores (simple sum, min score, max score, matcher weighting, and user weighting) were tested in this research. The EERs of the best unimodal fingerprint and the face recognition systems were 2.16% and 3.76%, respectively, while the max-score fusion approach using the quadric-linequadric technique over the normalized scores obtained an EER of 0.63%. Experiments conducted on a database of 100 users indicate that the application of min-max, z-score, and tanh normalization schemes followed by a simple sum of scores fusion method results in better recognition performance compared to other methods.

In the same year, Jain et al. studied the performance of different normalization techniques and fusion rules in the context of a multimodal biometric system based on the face, fingerprint and hand-geometry traits of a user at the score level [41]. Fingerprint matching was done using the minutiae features and the output of the fingerprint matcher was transformed into a similarity score. Eigenface coefficients were used to represent features of the face image. The Euclidean distance between the eigenface coefficients of the template and that of the input image was used as the matching score. The hand-geometry images were represented by a 14-dimensional feature vector and the matching score was computed as the Euclidean distance between the input feature vector and the template feature vector. Seven score normalization techniques (simple distance-t-similarity transformation with no change in scale, min–max normalization, z-score normalization, median-normalization, double-sigmoid normalization, tanh normalization, and Parzen normalization) and three fusion techniques on the normalized scores (simple sum rule, max rule, and min rule) were evaluated in this research. All fusion approaches outperform the unimodal approaches except the median-normalization. For instance, the fingerprint approach obtained a GAR of 83.6% at a FAR of 0.1%, while the multimodal approach obtained a GAR of 98.6% at a FAR of 0.1% when the z-score normalization and the sum rule were applied. The researchers observed that the tanh and min–max normalization techniques outperformed other techniques at low FARs, while the z-score normalization performs slightly better than the other techniques at higher FARs.

Xiuquin [186] proposes a multimodal biometric system using face and ear at feature level. Kernel discriminant analysis is employed as feature extraction method to obtain the features of face and ear independently and then concatenate the two feature vectors to form a single feature vector. Rattani et al. [151] proposed a multimodal biometric system of iris and face in which Scale Invariant Feature Transform (SIFT) features of individual modalities are extracted and concatenated to form the fused feature space.

From the previous review, we can conclude that several multimodal biometric systems with various methods and strategies have been proposed over the last few years to accomplish higher accuracy performance. In this context, we have also observed that, so far most of addressed techniques are based on the post-mapping fusion, that is, in decision and score matching levels of fusion. Only limited work is reported on feature level fusion of multimodal biometric system.

We have noticed also that the majority of the work reported on feature level fusion is related to multimodal biometric system using face and palmprint. Feng et al. [43] proposed the feature level fusion of face and palmprint in which PCA and ICA are used for feature extraction. Yao et al. [189] have proposed a multimodal biometric system using face and palmprint at feature level. In their research, Gabor features of face and palmprints are obtained individually. Extracted Gabor features are then analysed using linear projection scheme such as PCA to obtain the dominant principal components of face and palmprint separately. Finally, feature level fusion is carried out by concatenating the dominant principal components of face and palmprint to form a fused feature space. Jing et al. [77] employed Gabor transform for feature extraction and then Gabor features are concatenated to form fused feature vector. Then, to reduce the dimensionality of fused feature vector, nonlinear transformation techniques such as Kernel Discriminant Common Vectors are employed.

5.2 Feature Level Fusion

Fusion at the feature level is relatively an understudied problem [154]. Fusion at this level can be applied to the extracted features from the same modality or several multimodalities. In this work, we limit ourselves to iris and online signature cues. To the best of our knowledge, there is no reported research work combined iris and online signature.

5.2.1 Iris and Online Signature Fusion

The main reason behind the selection of iris and online signature as biometric features for building a multimodal biometric system stems from their strength points. The complex texture of the iris is unique and valuable source of personal recognition. The performance of currently deployed iris-based recognition systems is promising and encourages further research in the direction of large-scale identification systems based on iris information. Moreover, each iris possesses unique characteristics, and similar to fingerprints, even the irises of the eyes of identical twins are different [32]. It is extremely difficult to surgically alter the texture of the iris. Even though, early iris-based recognition approaches required significant user participation and were expensive, the current approaches have become more user-friendly and cost-effective [19].

For a long time, the way a person signs his or her name is known to be a distinguishing aspect of that individual. A handwritten signature is a behavioral biometric that change over a period of time and are influenced by physical and emotional conditions of the signatories. Signature has been widely accepted as a means of legal and commercial transactions identity authentication. Even though, hypothetically, no person write his/her signature exactly the same each time, in practice, it is very difficult to forge the dynamic information, such as speed, pen-up movement and pressure.

The main drawback of biometrics when compared with conventional authentication techniques is that many biometrics can be copied or forged. Whereas it is always possible to obtain another key or a new password, it is not possible to replace any biometric data [73]. However, signature is an exception, as users can be asked to change their signature if needed. A brief comparison between iris and signature is provided in Table 5.1 based on the perception of the authors of [170].

Table 5.1 Comparison between iris and signature biometric characteristics [170], H: high, M: medium, L: low

Biometrics	Security			Convenience		
	Required security level	Accuracy	Long-term stability	User acceptance	Ease of Use	Cost
Iris	H	H	H	M	M	H
Signature	M	H	M	H	H	L

5.2.2 Obstacles in Feature Fusion Scheme

It is believed that feature set contains richer information about the raw biometric data. Thus, integration at this level is of fusion is expected to act better in comparison with fusion at the score level and decision level [154,174]. Nevertheless, fusion at this level is a challenging problem due to a variety of reasons. Including that most feature sets gathered from multiple modalities are incompatible, such as in the case of combining fingerprint minutiae and eigenface coefficients [9]. Moreover, concatenating several feature vectors will lead to construct a very large feature vector or what is called the curse of dimensionality. This definitely increases the computational and storage resources demands. As Kludas et al. pointed out that a significantly more complex classifier design might be needed to operate on the concatenated data set at the feature level space [92]. Furthermore, poor feature representation, which mostly contains noisy or redundant data may sharply reduce the classification accuracy [41].

The problem of dimensionality reduction can be overcome by either performing feature transformation or feature selection. Feature selection, also known as feature reduction, attribute selection or variable subset selection, is the technique of selecting a subset of relevant features for building robust learning models [43]. By removing most irrelevant and redundant features from the data, feature selection helps improve the performance of learning model [5]. Assuming an original feature set of *n* features, the objective of feature selection is to identify the most informative subset of *m* features (*m* < *n*). Common feature selection approaches, such as sequential forward selection (SFS), sequential backward selection (SBS), sequential floating forward selection (SFFS), genetic algorithms (GA) have been applied successfully to several optimization tasks [53].

Feature transformation, on the other hand, represents the feature vector in another vector space to improve the representative-ness of the data. Moreover, only the significant "eigenvectors" are kept, inducing a subsequent reduction of dimension in the representation of the data. Finding such projection space requires a training phase on an adequate database. PCA, Linear Discriminant Analysis (LDA) and ICA [192] are three main linear techniques used for data reduction and feature transformation. Whereas, kernel PCA (KPCA) has been widely studied and applied in extracting nonlinear structures in data [186].

As particle swarm optimization (PSO) has been shown to be very efficient in optimizing the feature selection process in large scale application problems [70,83,151] we decided to deploy the binary particle swarm optimization (BPSO) algorithm to perform feature selection. Therefore, implementing BPSO in biometric feature fusion problem of high dimension is another novelty of this thesis.

Next section is devoted to the presentation of the PSO algorithm and to its implementation in the context of this thesis. The PSO algorithms for continuous and the BPSO are described, BPSO parameters and recommended settings are also discussed in detail.

5.3 Feature selection using PSO

PSO is an nature-inspired, evolutionary, population-based optimization algorithm whose goal is to minimize or to maximize an objective function $f: S \rightarrow \Re$. In this thesis, minimization problems are assumed, which means that the goal is to find a solution $x^* \in S$ such that $\forall x \in S: f(x^*) \leq f(x)$. A solution x^* that satisfies this condition is called a *global minimum* of *f*. If there exists an $\varepsilon > 0$ such that $\forall x$ *with* $\|x - x^*\| < \varepsilon : f(x^*) \leq f(x)$, the solution x^* is called a *local minimum*.

The PSO algorithm was developed by Kennedy and Eberhart in 1995 [83]. A detailed description with a lot of background information can be found in their textbook Swarm Intelligence [84].

5.3.1 Flocks, herds and schools

The main idea of swarm intelligence algorithm developed from inspiration of the collective intelligence of animal's societies that don't have any leader in their group or swarm, such as birds, ants, fish and termites. Their collective behaviours emerge from interactions among individuals, in a process known as self-organisation. Habitually, a flock of animals that have no leaders will find food by randomly following one of the members of the group with the closest position to the food source. The flocks achieve their best condition simultaneously through communication among members who already have a better location. Animal which has a better location will inform it to its flocks and the others will move simultaneously to that position. This process will be repeated until the best positions or a food source discovered [84]. Their collective behaviours emerge from interactions among individuals, in a process known as self-organisation. This collaborative behaviour among social animals exhibits a remarkable degree of intelligence. Each individual may not be intelligent by itself, but together they perform complex collaborative behaviours [17].

In PSO, each particle makes use of its own memory and knowledge gained by the swarm as a whole to find the best solution. Each potential solution is considered as a particle with a certain velocity, and ''*flies*'' through the problem space. Each particle adjusts its flight towards the target according to its own flying and its companions' flying experiences [172]. Hence, the particle swarms find optimal path towards destination through the interaction of individuals in a population of particles.

Thus, PSO has been successfully applied to a wide range of difficult combinatorial optimization applications [70]. PSO proved to be both effective and efficient in reducing feature dimension and removing irrelevant features.

5.3.2 Principles of PSO

In PSO, every possible candidate solution can be considered a *particle* in the search space. Each particle p_i makes use of its own memory and knowledge gained by the swarm as a whole to find the best solution. With the purpose of discovering the optimal solution, each particle adjusts its searching direction according to two factors, its own best previous experience (*pbest*) and the best experience of its companions flying experience (*gbest*). Shi and Eberhart [168] called pbest the cognition component, and gbest the social component. Each particle is moving around the n-dimensional search space S with objective function $f : S \subseteq \Re^n \to \Re$. Each particle has a position $x_{i,t}$ (t represents the iteration counter), a fitness function $f(x_{i,t})$ and ''flies'' through the problem space with a velocity $v_{i,t}$. A new position $z_1 \in S$ is called better than $z_2 \in S$ *iff* $f(z_1) < f(z_2)$.

Particles evolve simultaneously based on knowledge shared with neighboring particles; they make use of their own memory and knowledge gained by the swarm as a whole to find the best solution. The best search space position particle i has visited until iteration t is its previous experience *pbest*. To each particle, a subset of all particles is assigned as its neighbourhood. The best previous experience of all neighbours of particle i is called g*best*. Each particle additionally keeps a fraction of its old velocity, which results in the following update equations for particle swarm optimization [83].

$$v_{pd}^{new} = \omega * v_{pd}^{old} + C_1 * rand_1() * (pbest_{pd} - x_{pd}^{old}) + C_2 * rand_2() * (gbest_{d_d} - x_{pd}^{old}) \quad 5.1$$

$$x_{pd}^{new} = x_{pd}^{old} + v_{pd}^{new} \quad 5.2$$

In Equation 5.1, the first part is the previous flying velocity of the particle; while the second part represents the ‘‘*cognition*” part, which is the private thinking of the particle itself, where C1 is the individual factor. The third part is the ‘‘*social*” part, which represents the collaboration amongst the particles, where C2 is the societal factor [173].

The acceleration coefficients (C1) and (C2) are constants (also known as learning factors) represent the weighting of the stochastic acceleration terms that pull each particle toward the pbest and gbest positions. Therefore, the adjustment of these acceleration coefficients changes the amount of ‘tension’ in the system. Small values allow particles to travel far from target regions before being tugged back. In contrast, high values result in sudden movement toward, or past, target regions. In the original algorithm, the value of (C1 + C2) is usually limited to 4 [83].

Particles’ velocities are restricted to a maximum velocity, V_{max}. If V_{max} is too small, particles in this case may not travel around beyond local regions. They could become trapped in local optima. In contrast, if V_{max} is too high particles might fly past good solutions. According to Equation 5.1, the particle’s new velocity is calculated according to its previous velocity and the distances of its current position from its own best experience and the group’s best experience. Afterwards, the particle flies toward a new position according to Equation 5.2. The performance of each particle is measured according to a pre-defined fitness function which is related to the problem concerned [183]. The PSO algorithm is usually terminated either when a maximal number of generations is reached or when the best particle position of the entire swarm cannot be improved further after a sufficiently number of iterations.

5.3.3 Binary Particle Swarm Optimization

PSO was initially developed for a space of continuous values and it consequently, poses several problems for spaces of discrete values. Kennedy and Eberhart [85] presented a discrete binary version of PSO method (BPSO) for discrete optimization problems.

In BPSO, particles uses binary string to represent its position in form by $X_p=\{x_{p1},x_{p2},..., x_{pd}\}$ which is randomly generated. As each bit in the string represents a feature, value ‘1’ means that the corresponding feature is selected while ‘0’ means that it is not selected. The velocity of each particle is represented by

$V_p=\{v_{p1},v_{p2},...,v_{pd}\}$, where p is the number of particles, and d is the number of features of a given dataset. The initial velocities in particles are probabilities constrained to the interval [0.0–1.0]. In BPSO, using the knowledge of pbest and gbest, the features of the pbest and gbest particles can be obtained with regard to their position and velocity. Each particle is updated according to the following equations [85]:

$$v_{pd}^{new} = w * v_{pd}^{old} + C_1 * rand_1() * (pbest_{pd} - x_{pd}^{old}) + C_2 * rand_2() * (gbest_d - x_{pd}^{old}) \quad 5.3$$

$$if\ (V_{pd}^{new}) \notin (V_{\min}, V_{\max}) \quad \text{then} \quad V_{pd}^{new} = \max\ (\min(V_{\max}, V_{pd}^{new}), V_{\min}) \quad 5.4$$

$$S(V_{pd}^{new}) = \frac{1}{1+e^{-v_{pd}^{new}}} \quad 5.5$$

$$x_{pd}^{new} = \begin{cases} 1 & if\ (r_3 < S(V_{pd}^{new})) \\ 0 & otherwise \end{cases} \quad 5.6$$

In Equation 5.3, w is the inertia weight, C1 and C2 are acceleration parameters, and *rand*, *rand1* and *rand2* are three independent random numbers in the range [0, 1]. Velocity V_{pd}^{new} is the updated particle and V_{pd}^{old} is the velocity of the particle before being updated, x_{pd}^{old} is the current particle position and x_{pd}^{new} is the updated particle position.

In Equation 5.4, particle velocities of each dimension are limited to within $[V_{\min}, V_{\max}]^D$. If the velocity of that dimension to exceed $V_{\max}$ as a result of the summation of the two accelerations then the velocity of that dimension will be limited to $V_{\max}$. In Equations 5.5 and 5.6, the updated positions of the particles are calculated by the function $S(V_{pd}^{new})$, where V_{pd}^{new} is the updated velocity value.

If the function $S(V_{pd}^{new})$ is larger than r_3, which is the randomly produced disorder number that is within the range of [0.0–1.0], then its position of the particle x_{pd}^{new} will be updated to 1, which means that this feature is selected as a required feature for the next update. Otherwise, the x_{pd}^{new} will be assigned to 0, which means that this feature is no longer required for the next update cycle. A flowchart of BPSO is shown in Figure 5.1.

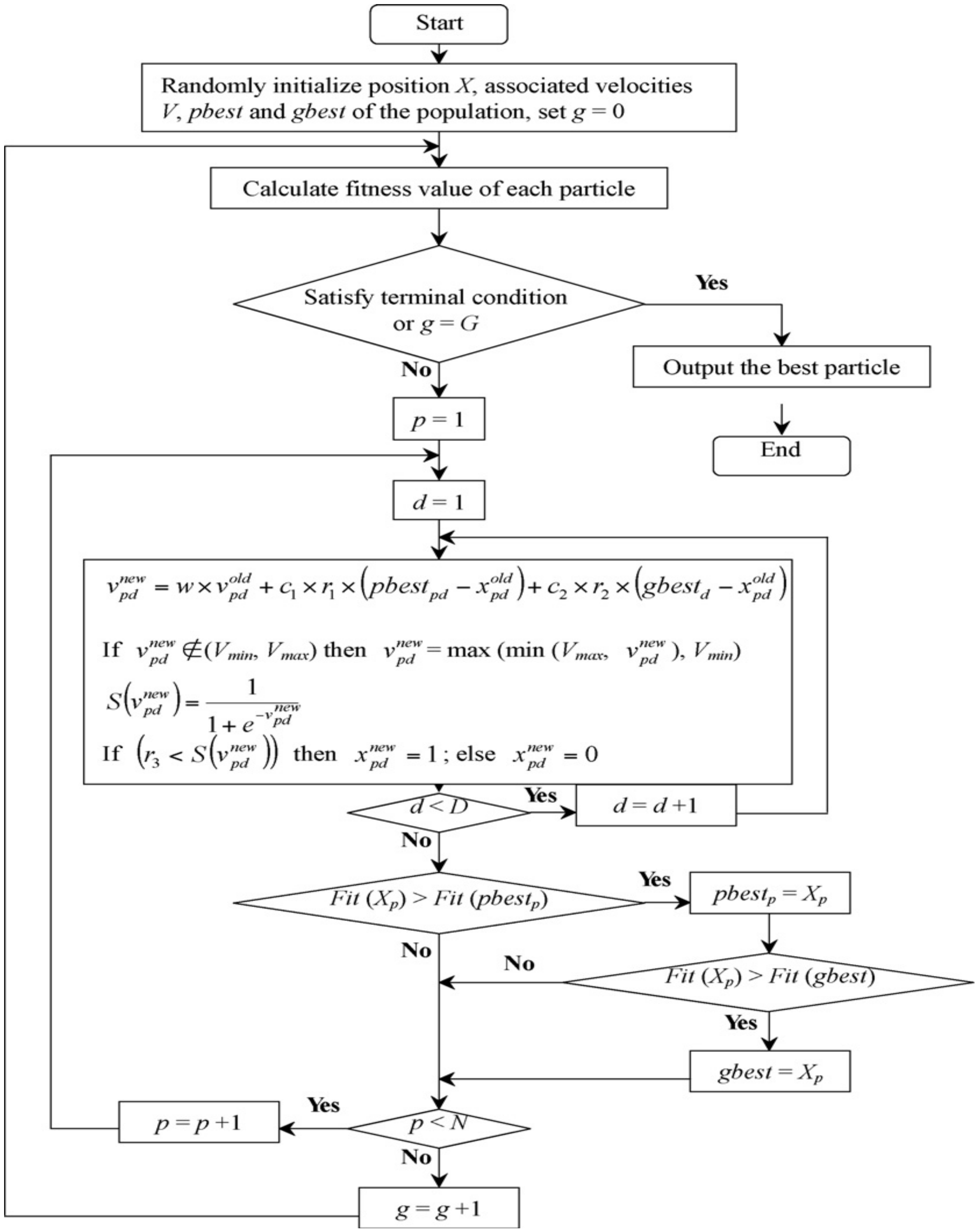


Figure 5.1 The BPSO flow chart, *adapted from* [27]

p: number of particles, d: number of features, D: total number of features, g: number of generations, G: maximum number of generations, and N: population size.

5.3.4 BPSO Implementation Details

Feature selection is a crucial process typically carried out to select an optimal subset of features and remove the redundant and irrelevant features that cause classification degradation. Numerous feature selection algorithms can be used to perform feature selection of multimodal biometrics features.

In this chapter, we select the binary PSO to perform a selection of iris and online signature combined features for the many of reasons [85].

- Firstly, feature selection techniques typically involve searching large dimensional vector space and PSO has demonstrated that its performance is insensitive to the population size [168]. Therefore, it was successfully applied in numerous applications such as dynamical systems, operations research, bioinformatics, medical informatics, noisy and dynamic environments [138].

- Secondly, PSO requires only simple mathematical operations compared with complex evolution operators such as crossover and mutation used in Genetic Algorithms. Hence, PSO is conceptually simple in terms of both memory requirements and speed.

- Lastly, each particle swarm has a memory remembering the best position of the search space that has ever been visited. Therefore, the knowledge of good solutions is retained by all particles [27].

The selection of PSO parameters can have a considerable impact on the performance of optimization [85]. Therefore, selecting PSO parameters that yield good performance has been the subject of a lot research [27]. In this section, we describe several PSO parameters such as fitness function, acceleration constant, inertia weight and velocity limitation which need to be estimated before conducting experiments.

1. Fitness function

The PSO implementation relies on the appropriate formulation of the fitness function. The main objective of the closed identification fitness function is to maximize the recognition rate.

Given the test sample, we compute its distance against all the samples in the reference dataset to obtain the match scores. Then, we select the sample from the reference dataset with the lowest distance value and we check whether it belongs to the same class as the testing sample. We will repeat this for all testing samples and count the number of success and failures. In every iteration, each particle is evaluated, and a value of goodness or fitness of a given trail solution is returned by a fitness function. The fitness function F evaluates the quality of evolved particles in terms of their ability to maximize the class separation term indicated by the scatter index among the different classes [3]. Let w_1, w_2,..,w_L and N_1, N_2,...,N_L denote the classes and number of features within each class, respectively. Let M_1, M_2,...,M_L and M_o be the means of corresponding classes and the grand mean in the feature space, M_i can be calculated as:

$$M_i = \frac{1}{N}\sum_{j=1}^{N} W_j^i \quad , \quad i = 1,2,...,L \qquad 5.7$$

Where $W_j^i \quad , \quad i = 1,2,...,L$ represents the sample features from class w_i and the grand mean M_o is

$$M_o = \frac{1}{N}\sum_{j=1}^{L} N_i M_i \qquad 5.8$$

Where 'N' is the total dimension of the feature set. Thus, we define the fitness function F as follows:

$$F = \sqrt{\sum_{i=1}^{L} (M_i - M_o)^t (M_i - M_o)} \qquad 5.9$$

2. Velocity limitation Vmax

The velocity limit V_{max} plays an important role as it in the binary version of PSO, the value of V_{max} limits the probability that bit x_{id} will take on a value of 1 or 0 and consequently the use of high V_{max} value in BPSO will decrease the mutation rate [85]. In this thesis, we have tried several values of V_{max} and at last set V_{max} to 2, as we noticed it allows the particle to reach an optimum solution.

3. Inertia weight and acceleration constant

The weight of inertia is an essential variable in the BPSO algorithm as it affords the particles with a degree of memory capability. Many experimental studies found that inertia weight ‘‘w’’ in the range of [0.8, 1.2] leads to a good performance [85]. Therefore in this chapter, we initially set ‘‘w’’ to 0.6 in all iterations.

Although the rate of acceleration constants C1 and C2 are not so significant in the convergence of PSO, carefully chosen value may lead to faster convergence. In our experiments, we varied the value of C1 and C2 from 0 to 2 and finally chose C1=2 and C2=2.

4. Population size

The population size of PSO influences the performance and the computation cost. In our experiments, we experimentally varied the size of the population from 20 to 35 and finally, we fixed the population size as 30.
The parameters used for the BPSO are summarized in Table 5.2.

Table 5.2 Summary of BPSO parameters

Parameters	Values
Population size	30
Maximum Number of iterations (G)	100
Velocity limitation (V_{max})	2
Inertia weight (w)	0.6
Acceleration constant (C1 and C2)	C1=C2=2

5.4 Feature Level Fusion of Iris and Signature

As described in former chapters, we have extracted the features of iris and online signature separately. For the iris modality, we have applied the 2D-DTCWT on the normalized iris images and obtain the real parts of the complex coefficients at several resolution scales to form a feature vector that represents the iris image. Whereas, for online signature modality, we obtained a selection of 31 global functions to represent signature dynamics. Figure 5.2 shows the proposed block diagram of feature level fusion of iris and online signature before deploying BPSO.

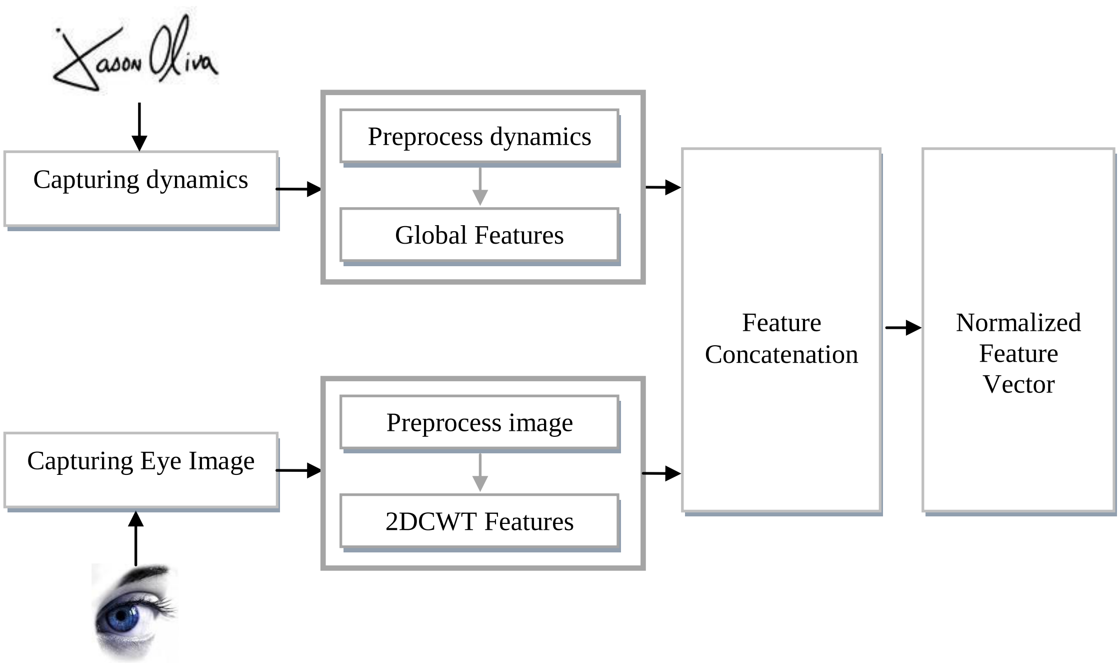


Figure 5.2 Schematic for proposed multimodal identification scheme based on the fusion of iris and online signature

Afterwards, we obtain a joint feature vector by vertically concatenating the columns of the iris and online signature features. We repeat this for all the subjects in the database in order to obtain a complete fused set for the entire database as explained as follows.

Let S_{Iris} = [S_{1Iris}, S_{2Iris},…, S_{NIris}] represents the iris extracted features and $S_{Signature}$=[$S_{1Signature}$, $S_{2Signature}$,…, $S_{NSignature}$] represents the online signature features.

We vertically concatenate S_{Iris} and $S_{Signature}$ to obtain the fused feature vector $X_{FusedFeatures}$= [S_{1Iris}, S_{2Iris},…, S_{NIris}, $S_{1Signature}$, $S_{2Signature}$,…, $S_{NSignature}$] and we repeat this for all the subjects to obtain a new fused set $X_{FusedFeatures}$.

As the fused feature values of vectors of signature and iris exhibit significant variations both in their range and distribution, feature vector normalization is carried out. The objective behind feature normalization (also called *range-normalization*) is to modify the location (*mean*) and scale (*variance*) of the features values and to independently normalize each feature component to the range between 0 and 1 as follows [44].

$$\overline{X}_{FusedFeatures} = \frac{X_{FusedFeatures} - \mu_{FusedFeatures}}{\delta_{FusedFeatures}} \qquad 5.10$$

Where $\mu_{FusedFeatures}$ and $\delta_{FusedFeatures}$ indicates the mean and variance value of $X_{FusedFeatures}$. Finally we obtain the normalized feature vector set $\overline{X}_{FusedFeatures}$. Figure 5.3 illustrates the proposed PSO-based feature selection algorithm.

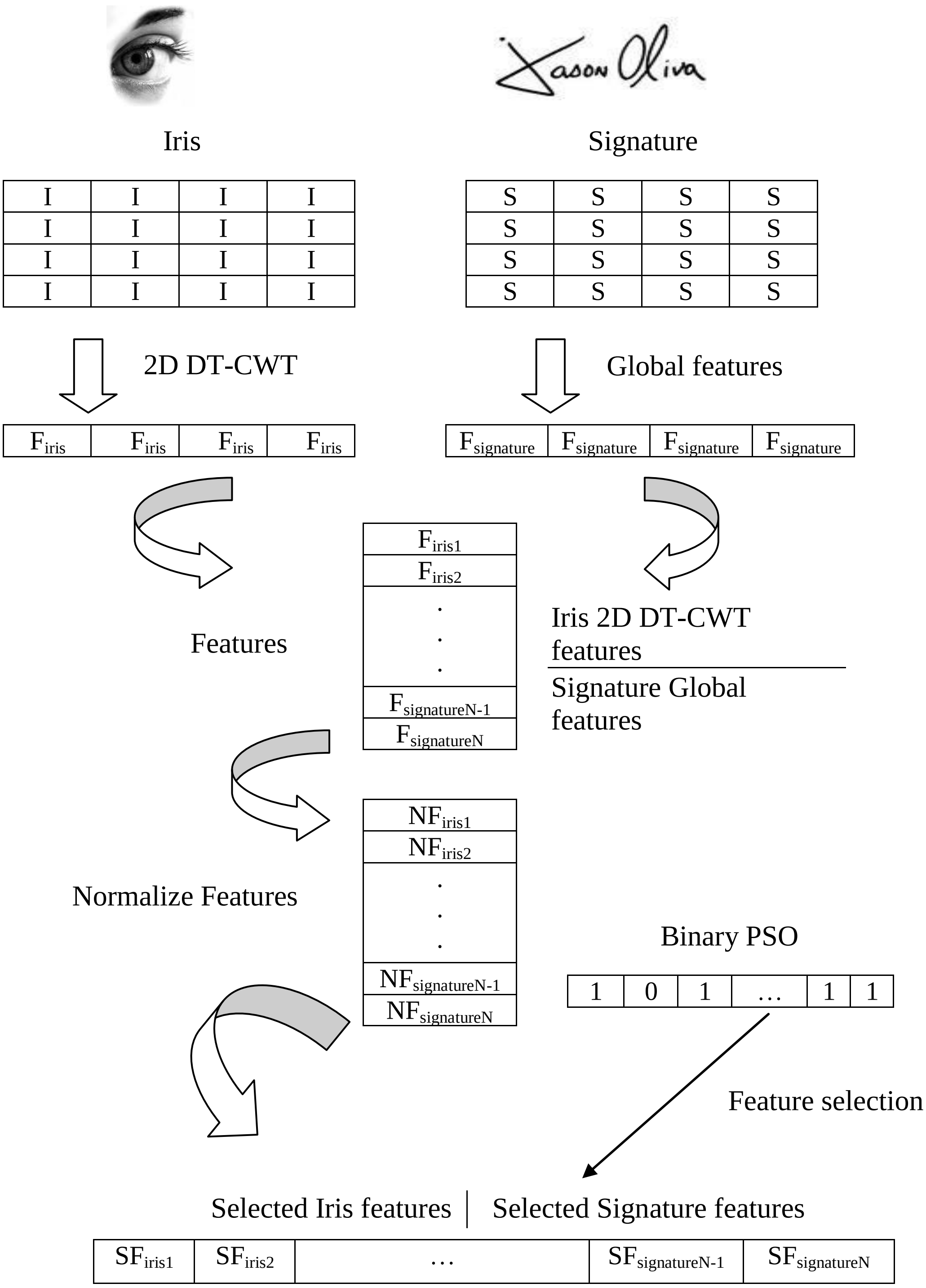


Figure 5.3 BPSO Proposed scheme of feature fusion selection (scheme I)

5.4.1 Suggested feature level scenarios

As previously stated, we plan to design a simple iris-signature multi biometrics system based on feature level fusion. In real-world application, the feature set is generally large in terms of dimensionality. Usually, the resulting feature vector may be noisy and contain irrelevant or redundant information about the target classes. This may possibly degrade the performance of the classifiers. Furthermore, large feature vector also increases the storage cost and requires more computation time to process it [80].

Feature selection in this case, is crucial to select an "optimized" subset of features from the original feature set based on certain objective function. Overall, feature selection removes redundant or irrelevant data while retaining classification accuracy.

Scheme I

In feature fusion **scheme I**, the proposed scheme is based on the idea of applying the PSO on the normalized companied features $\overline{X}_{FusedFeatures}$ in order to select the most dominant features from the fused feature space. We will now describe in detail the steps needed to implement this scheme:

Step 1: Apply the 2DT-CWT on the extracted iris images to obtain S_{Iris} wavelet coefficients. *Step 2: Extract the global features from the dynamic signatures $S_{Signature}$.* *Step 3: Vertically concatenate S_{Iris} and $S_{Signature}$ to obtain the fused features vector $X_{FusedFeatures}$.* *Step 4: Normalize the fused feature using min-max normalization to obtain $\overline{X}_{FusedFeatures}$.* *Step 5: Randomly initialize the PSO particles with binary values (0 and 1).* *Step 6: Carry out the feature selection by considering the value of the bit in the particle. More precisely, if bit value is 1, select the corresponding feature from $\overline{X}_{FusedFeatures}$.This way we construct a new feature vector.*

Scheme II

Whereas, **Scheme II** starts first with performing the PCA to reduce the size of S_{Iris} and then further reducing the dimension of the fused features $\overline{X}_{FusedFeatures}$ using the BPSO before performing the matching step. Here is a description of the steps involved in scheme II:

Step 1: Apply the 2DT-CWT on the extracted iris images to obtain S_{Iris} wavelet coefficients.
Step 2: Extract the global features from the dynamic signatures $S_{Signature}$.
Step 3: Apply PCA to the iris feature vector S_{Iris} to obtain S_{IrisR}.
Step 4: Vertically concatenate S_{IrisR} and $S_{Signature}$ to obtain the fused features vector $X_{FusedFeatures}$.
Step 5: Normalize the fused feature using min-max normalization to obtain $\overline{X}_{FusedFeatures}$.
Step 6: Randomly initialize the PSO particles with binary values (0 and 1).
Step 7: Carry out the feature selection by considering the value of the bit in the particle. More precisely, if bit value is 1, select the corresponding feature from $\overline{X}_{FusedFeatures}$ .This way we construct a new feature vector $\overline{X}_{NewFusedFeatures}$.
Step 8: Apply PCA to the new feature vector $\overline{X}_{NewFusedFeatures}$.

Scheme III

Scheme III starts with applying the BPSO to the normalized fused feature vector $\overline{X}_{FusedFeatures}$ and then it applies the PCA on the remaining fused features. The procedure is as following:

Step 1: Apply the 2DT-CWT on the extracted iris images to obtain S_{Iris} wavelet coefficients.
Step 2: Extract the global features from the dynamic signatures $S_{Signature}$.
Step 3: Vertically concatenate S_{Iris} and $S_{Signature}$ to obtain the fused features vector $X_{FusedFeatures}$.
Step 4: Normalize the fused feature using min-max normalization to obtain $\overline{X}_{FusedFeatures}$.
Step 5: Randomly initialize the PSO particles with binary values (0 and 1).
Step 6: Carry out the feature selection by considering the value of the bit in the particle. More precisely, if bit value is 1, select the corresponding feature from $\overline{X}_{FusedFeatures}$ .This way we construct a new feature vector $\overline{X}_{NewFusedFeatures}$.
Step 7: Apply PCA to the resulting feature vector $\overline{X}_{NewFusedFeatures}$.

Scheme IV

Scheme IV is quite similar to scheme III; yet, it starts with applying the PCA to the normalized companied feature vector $\overline{X}_{FusedFeatures}$ and then it applies the BPSO on the rest of the fused features. Here is the proposed scheme in detail:

Step 1: Apply the 2DT-CWT on the extracted iris images to obtain S_{Iris} wavelet coefficients.
Step 2: Extract the global features from the dynamic signatures $S_{Signature}$.
Step 3: Vertically concatenate S_{Iris} and $S_{Signature}$ to obtain the fused features vector $X_{FusedFeatures}$.
Step 4: Normalize the fused feature using min-max normalization to obtain $\overline{X}_{FusedFeatures}$.
Step 5: Apply PCA to the resulting feature vector $\overline{X}_{FusedFeatures}$ to obtain $\overline{X}_{NewFusedFeatures}$.
Step 6: Randomly initialize the PSO particles with binary values (0 and 1).
Step 7: Carry out the feature selection by considering the value of the bit in the particle. More precisely, if bit value is 1, select the corresponding feature from $\overline{X}_{NewFusedFeatures}$.

Scheme V

Scheme V is quite similar to schemes III and IV; yet, it starts with applying the BPSO to the normalized fused feature vector $\overline{X}_{FusedFeatures}$ and then it applies the CFS on the remaining fused features. The procedure is as following:

Step 1: Apply the 2DT-CWT on the extracted iris images to obtain S_{Iris} wavelet coefficients.
Step 2: Extract the global features from the dynamic signatures $S_{Signature}$.
Step 3: Vertically concatenate S_{Iris} and $S_{Signature}$ to obtain the fused features vector $X_{FusedFeatures}$.
Step 4: Normalize the fused feature using min-max normalization to obtain $\overline{X}_{FusedFeatures}$.
Step 5: Randomly initialize the PSO particles with binary values (0 and 1).
Step 6: Carry out the feature selection by considering the value of the bit in the particle. More precisely, if bit value is 1, select the corresponding feature from $\overline{X}_{FusedFeatures}$ .This way we construct a new feature vector $\overline{X}_{NewFusedFeatures}$.
Step 7: Apply CFS to the resulting feature vector $\overline{X}_{NewFusedFeatures}$.

Finally, the decision about accept/reject in all the schemes is evaluated using number of supervised learning classifiers. For comparison purpose, we implemented and evaluated three classifiers, namely: Naïve Bayes, k-NN and SVM.

Figures 5.4-5.8 show the block diagram of the proposed feature fusion, where all the techniques start with acquiring the features of iris and online signatures separately.

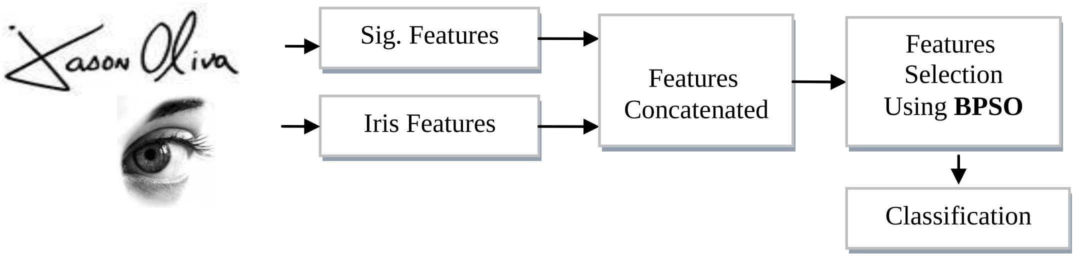


Figure 5.4 BPSO Proposed scheme of feature fusion (Scheme I)

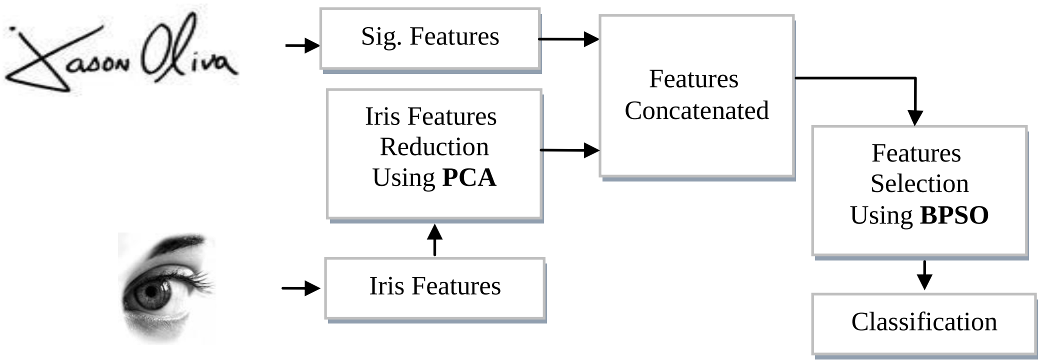


Figure 5.5 PCA-BPSO Proposed scheme of feature fusion (Scheme II)

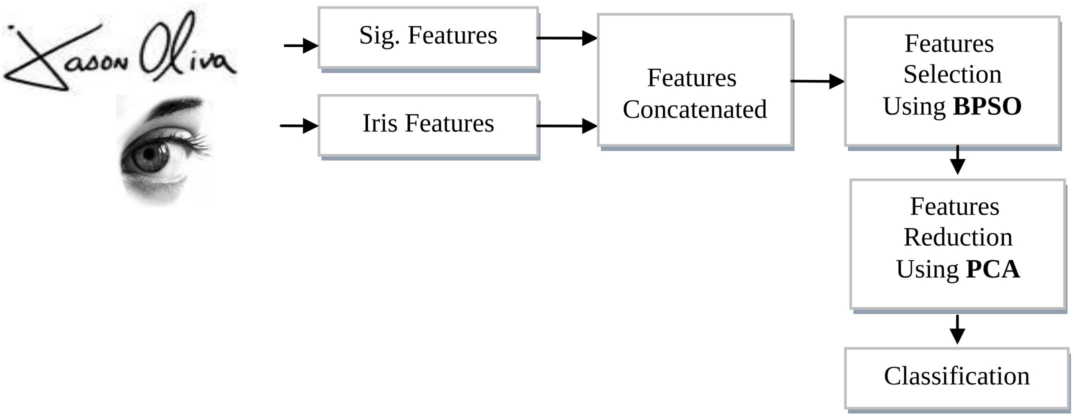


Figure 5.6 BPSO-PCA Proposed scheme of feature fusion (Scheme III)

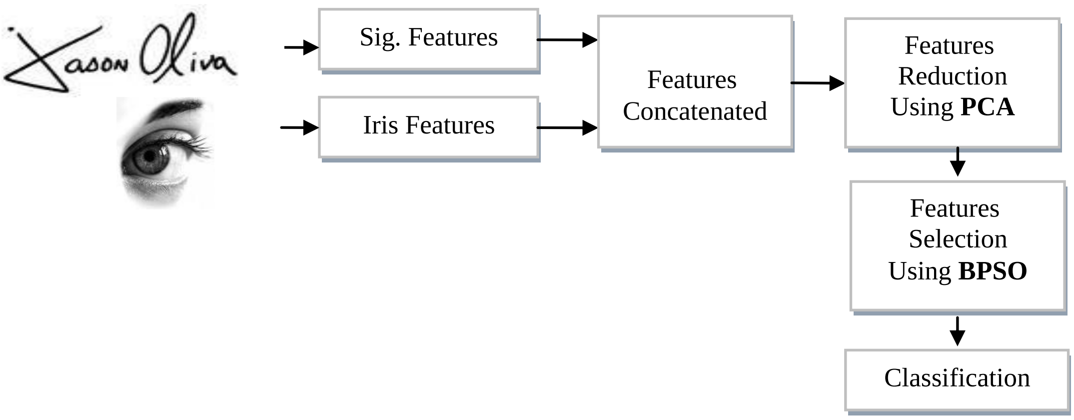


Figure 5.7 PCA-BPSO Proposed scheme of feature fusion (Scheme IV)

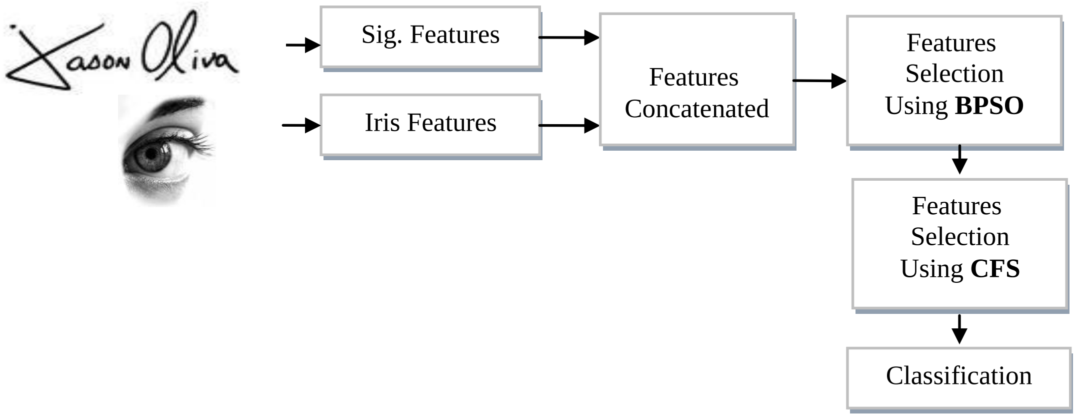


Figure 5.8 BPSO-CFS Proposed scheme of feature fusion (Scheme V)

5.5 Experimental Results

This section describes the experimental setup, including database and the assessment protocol that we have built in order to evaluate the proposed feature level fusion schemes.

5.5.1 On the Use of Chimeric Users in Multimodal Biometric

The first difficulty we are facing when working on multi-biometrics is the lack of real-user databases. In order to evaluate the performance of a multimodal system based on iris and online signature modalities, it is essential to have a database that contains data of the two modalities. Unfortunately, as far our knowledge is concerned, there is no public multimodal real-user database which combines online signature and iris modalities of the same individuals available. However, there exist few well established datasets for iris images, thus implying the combination of biometric modalities from different databases. Since both databases do not necessarily contain the same users, such combination results in the creation of virtual multimodal dataset, or so-called chimeric users.

Creating such chimeric users has lately been widely accepted and reasonable practice in the field of multimodal biometrics research as a way to overcome the problem of shortage of actual multimodal biometric databases [154]. An investigation into the using of chimeric users to construct fusion classifiers in biometric authentication tasks was reported in [72] with the conclusion that a fusion operator derived from multiple chimeric-user databases does not enhance nor degrade the generalization performance (on real users) with respect to training it on real users databases.

5.5.2 The Chimeric Database

We have created a '*virtual multimodal database*' by aggregating two different database using online signature and iris modalities coming from two different databases. A user from the online signature dataset is randomly associated with a user from the iris dataset, creating a virtual user with online signature and iris samples.

For online signature modality, we chose the database we have gathered in Chapter Three. The signature database contains 2160 signatures of 108 volunteers with 20 images per each class taken from two sessions, and each session was taken with an interval of several weeks. From this database, we selected 7 signature scripts from 108 different users. For the iris modality, we chose the CASIA eye image database version 1.0 [21]. The CASIA database contains 756 frontal iris images of 108 classes with 7 images per each class taken from two sessions, and each session were taken with an interval of one month. In building our multimodal biometric database of online signature and iris, each virtual subject was associated with 7 randomly samples of iris and online signature from two subjects in the aforementioned databases. Thus, the resulting virtual multimodal biometric database consists of 108 subjects, so that each subject has seven samples.

5.5.3 Results and discussion

As mentioned earlier, the first set of experiments (Scheme I) is based on applying BPSO after fusing the features of the iris and signature. Whereas, the second, the third and the fourth feature fusion experiments (i.e. schemes II, III and IV), study the effect of further reducing the same set of reduced set of features using PCA prior/subsequent to classification. While the last scheme, Scheme IV, study the effect of reducing the obtained set of features using CFS after applying BPSO to the fused feature vector. Note that for the feature fusion schemes I_a ,II_a ,III_a , IV_a and V_a we have applied all the 31 extracted signature features, while for the feature fusion schemes I_b, II_b, III_b , IV_b and V_b we have applied the minimal reducted set of signature features using the rough set.

All experiments were carried out using 10-fold cross-validation to minimize the bias associated with the random sampling of the training. In 10-fold cross-validation the whole database is randomly partitioned into 10 mutually and approximately equally sized subsets. The classification task is carried out 10 times, each time using one distinct partition as the testing set and the remaining 9 partitions as the training set. Thus, 10 different test results exist for each training test configuration. The precision and recall is computed as the average of the total runs.

Table 5.3 shows the best classification rate and the number of features, together with the classifier applied in building unimodal approach. It is clear that the performance of the online signature unimodal system outperforms the iris unimodal model with GAR of 97.48% achieved with 31 features and a GAR of 95.11% with 9 features using the k-NN classifier. Whilst, the iris modality achieved a GAR of 92.86% with a feature vector of at the third level of 2D-DTCWT decomposition using the SVM Gaussian RBF kernel. The fact that the results of the online signature features are better than iris is due to that the fact that the iris preprocessing phase has trimmed a significant part of the iris with the purpose of eliminating the effect of the eye lid and eye lashes.

Table 5.3 Unimodal recognition rates

Method	Classifier	Number of features	Recognition Rate (%)
Iris *alone*	SVM-RBF	1040	92.86
	SVM- Polynomial	1040	92.46
	SVM-Linear	1040	92.32
	k-NN	1040	80.82
	Naïve Bayes	280	77.11
Online signature *alone*	k-NN	31	97.48
	k-NN	9	95.11
	Naïve Bayes	31	94.57
	Naïve Bayes	9	93.91

Tables 5.4 and 5.5 show the performance of feature fusion scheme I_a and scheme I_b. The tables present the performance of each classifier along with the number of features obtained after applying the BPSO. It is observed that, the best performance is noted for fusion scheme I_a was a GAR of 98.14% with the SVM-RBF kernel with a feature vector of 50, while the fusion scheme I_b scored the best classification rate of 93.78% using 45 features using the SVM-RBF kernel. Naïve bayes and k-NN classifiers recorded a GAR of 94.84 and 97.08 with 80 features, respectively.

We also observed that, Scheme I_a recorded better classification rates than Scheme I_b. We have noticed that the SVM has outperformed the other classifiers in most of the experiments.

It can be noticed from both tables that the best classification results was recorded when fusing the iris features of the 5^{th} level of 2D-DTCWT decomposition with the online signature features.

Table 5.4 BPSO Proposed scheme of feature fusion (scheme I_a)

Biometrics			Classifier Accuracy rate (%)					
Online signature	Iris	dim.	Naïve Bayes	k-NN	SVM			
					Gaussian	RBF	Polynomial	Linear
31	1040	515	84.78	93.25	93.65	94.44		94.44
31	280	155	91.93	95.37	95.50	95.02		95.63
31	80	50	94.84	97.08	98.14	97.22		98.01
31	24	24	94.04	95.76	97.88	96.82		97.22

Table 5.5 BPSO Proposed scheme of feature fusion (scheme I_b)

Biometrics			Classifier Accuracy rate (%)					
Online signature	Iris	dim.	Naïve Bayes	k-NN	SVM			
					Gaussian	RBF	Polynomial	Linear
9	1040	524	80.02	92.67	92.98	93.65		93.65
9	280	134	64.94	81.34	83.46	80.82		82.67
9	80	45	87.30	91.66	93.78	93.12		93.51
9	24	12	86.50	84.65	90.07	88.09		89.81

In the next set of experiments, the considered fusion method Scheme II, is applied based on performing the PCA first to reduce the dimensionality of iris features S_{Iris} to the half. Followed by combining the resulted feature set with the online signature feature set prior to further reducing the dimension of the combined features $\overline{X}_{FusedFeatures}$ using the BPSO before performing the matching step. Tables 5.6 and 5.7 present the results in terms of GARs again for all the possible feature/classifier combinations for iris and signature features, respectively.

We noticed that this fusion approach shows similar performance as compared with the previous scheme. Nevertheless, the best performance is noted with the fusion scheme II_a by SVM classifier with Linear kernel function which scored a GAR of 98.94% with 282 features. The highest GAR (95.76%) in Scheme II_b is observed in the case of combing the 9 signature features with the 6^{th} level of 2D-DTCWT decomposition using the SVM Gaussian RBF function. We also noticed that, Scheme II_a recorded better classification rates than Scheme II_b.

The best performance for this scheme in most of the experiments is achieved with SVM classifiers is used. However, the achieved performance of the Naïve Bayes classifier suggests that it may be most sensitive to irrelevant and redundant features.

Table 5.6 PCA-BPSO Proposed scheme of feature fusion (scheme II_a)

Biometrics			Classifier Accuracy rate (%)				
Online signature	Iris	dim.	Naïve Bayes	k-NN	SVM		
					Gaussian RBF	Polynomial	Linear
31	1040	282	27.64	52.51	98.54	98.54	98.94
31	280	76	94.70	94.97	97.88	97.61	97.75
31	80	41	94.44	97.35	98.14	97.88	98.14
31	24	22	94.04	95.76	97.22	98.54	98.67

Table 5.7 PCA-BPSO Proposed scheme of feature fusion (scheme II_b)

Biometrics			Classifier Accuracy rate (%)				
Online signature	Iris	dim.	Naïve Bayes	k-NN	SVM		
					Gaussian RBF	Polynomial	Linear
9	1040	272	74.73	20.76	51.19	51.32	52.11
9	280	76	86.77	89.02	94.70	92.72	92.72
9	80	21	82.67	84.65	86.50	84.78	84.78
9	24	16	94.97	95.1	95.76	95.63	95.63

As described previously, Scheme III was carried out by applying BPSO first to the fused feature space $\overline{X}_{FusedFeatures}$. We then carry out the PCA to look for an axis in the kernel space that highlights the difference between classes. The best performance is noted for fusion scheme III_a by the k-NN classifier with a GAR of 98.01% and 30 features. The highest GAR (96.29%) in Scheme III_b is observed in the case of combing the 9 signature features with the 5^{th} level of 2D-DTCWT iris coefficients using the k-NN classifier. Clearly the results suggest that applying feature reduction with PCA after BPSO did not enhance the performance. It is also observed that, Scheme III_a recorded better classification rates than Scheme III_b.

Table 5.8 BPSO-PCA Proposed scheme of feature fusion (scheme III_a)

Biometrics			Classifier Accuracy rate (%)					
Online signature	Iris	dim.	Naïve Bayes	k-NN	SVM			
					Gaussian	RBF	Polynomial	Linear
31	1040	257	61.37	81.21	64.94	89.55	94.57	
31	280	87	81.48	96.95	92.98	91.37	95.10	
31	80	30	88.88	98.01	96.82	92.85	97.35	
31	24	13	90.87	95.89	96.56	93.25	96.95	

Table 5.9 BPSO-PCA Proposed scheme of feature fusion (scheme III_b)

Biometrics			Classifier Accuracy rate (%)					
Online signature	Iris	dim.	Naïve Bayes	k-NN	SVM			
					Gaussian	RBF	Polynomial	Linear
9	1040	272	74.735	51.19	51.19	51.32	52.11	
9	280	72	80.95	96.29	91	86.37	92.72	
9	80	18	79.36	91.53	91.53	83.33	90.87	
9	24	7	75	79.76	82.86	55.42	82.27	

In the following set of experiments, Scheme IV, we started with applying the PCA to the combined features $\overline{X}_{FusedFeatures}$, and then we applied the BPSO on the remainders of the fused features. The best performance is noted for this fusion scheme was achieved by the SVM classifier with the Gaussian RBF function with a GAR of 98.48% with 37 features. The highest GAR (95.76%) in Scheme IV_b is observed in the case of combing the 9 signature features with the 5^{th} level of 2D-DTCWT decomposition using the Gaussian RBF kernel function. We also noticed that Scheme IV_a recorded better classification rates than Scheme IV_b. We observed that the SVM has outperformed the other classifiers in most of this experiment set.

Table 5.10 PCA-BPSO Proposed scheme of feature fusion (scheme IV_a)

Biometrics			Classifier Accuracy rate (%)				
			Naïve Bayes	k-NN	SVM		
Online signature	Iris	dim.			Gaussian RBF	Polynomial	Linear
31	1040	263	86.5	31.87	67.98	66.66	69.97
31	280	81	73.94	91.4	94.84	63.09	94.97
31	80	37	93.78	97.48	98.48	98.28	98.28
31	24	14	86.64	93.38	91	40.47	90.21

Table 5.11 PCA-BPSO Proposed scheme of feature fusion (scheme IV_b)

Biometrics			Classifier Accuracy rate (%)				
			Naïve Bayes	k-NN	SVM		
Online signature	Iris	dim.			Gaussian RBF	Polynomial	Linear
9	1040	281	73.54	26.98	56.87	61.5	63.49
9	280	67	71.42	90.34	91.13	36.11	91.13
9	80	29	85.31	93.91	95.37	93.38	93.78
9	24	12	90.97	86.9	90.6	88.75	86.24

In the last set of experiments, Scheme V, we start with applying the BPSO to the combined features $\overline{X}_{FusedFeatures}$, and then we apply the CFS on the remainders of the fused features. The best performance is noted for this fusion scheme by the k-NN classifier with a GAR of 98.94% from 18 features. The highest GAR (93.65%) in Scheme V_b is observed in the case of combing the 9 signature features with the 4^{th} level of 2D-DTCWT decomposition using the SVM Gaussian RBF function. We also noticed that Scheme V_a recorded better classification rates than Scheme V_b. We observed that the SVM has outperformed the other classifiers in most of this experiment set.

Table 5.12 BPSO-CFS Proposed scheme of feature fusion (scheme V_a)

Biometrics			Classifier Accuracy rate (%)				
					SVM		
Online signature	Iris	dim.	Naïve Bayes	k-NN	Gaussian RBF	Polynomial	Linear
31	1040	24	93.78	95.89	96.95	94.44	96.95
31	280	25	95.63	97.08	97.48	95.48	97.61
31	80	18	95.76	98.94	98.41	87.433	98.67
31	24	7	87.03	83.46	98.01	97.48	86.11

Table 5.13 BPSO-CFS Proposed scheme of feature fusion (scheme V_b)

Biometrics			Classifier Accuracy rate (%)				
					SVM		
Online signature	Iris	dim.	Naïve Bayes	k-NN	Gaussian RBF	Polynomial	Linear
9	1040	37	80.48	90.21	91.79	91.66	91.40
9	280	14	92.85	90.34	93.65	92.32	93.12
9	80	11	89.41	91.26	91.53	90.47	92.46
9	24	7	87.03	83.46	87.03	85.18	86.11

Table 5.14 shows the comparative recognition rates of the suggested feature selection schemes.

Table 5.14 Comparative recognition rates of the different feature selection schemes

Method	Classifier	Number of features	Recognition Rate (%)
Iris alone	SVM-RBF	1040	92.86
Online signature alone	k-NN	31	97.48
Online signature alone	k-NN	9	95.11
Feature Fusion-Scheme I_a	SVM-RBF	50	98.14
Feature Fusion-Scheme I_b	SVM-RBF	45	93.78
Feature Fusion-Scheme II_a	k-NN	30	98.01
Feature Fusion-Scheme II_b	k-NN	72	96.29
Feature Fusion-Scheme III_a	k-NN	30	98.48
Feature Fusion-Scheme III_b	k-NN	72	96.29
Feature Fusion-Scheme IV_a	SVM-RBF	37	98.48
Feature Fusion-Scheme IV_b	SVM-RBF	29	95.76
Feature Fusion-Scheme V_a	k-NN	18	98.94
Feature Fusion-Scheme V_b	SVM-RBF	14	93.65

A number of important outcomes of the experimental analysis can be observed by considering the results in all the tables shown above. From these results, it is clearly seen that the best performance is noted for fusion scheme V_a was a GAR of 98.94% with the k-NN classifier with a feature vector of size 18 which outperforms the online signature in terms of accuracy rate and size of feature vector.

The experimental results showed that the SVM classifier achieved the best performance which is closely followed by k-NN. The performance of the Naïve Bayes was the worst, this may be due to the large number of features that make the repeated portioning of data not easy. However, the performance of k-NN is also promising.

It can be observed that the performance of any of the feature level fusion methods is superior to that of iris modality alone. More importantly, the feature fusion schemes with the 31 online signature features showed a better performance as compared with the feature fusion schemes with the reducted set of online signature features. This clearly indicates that the number of online signature features plays a significant role in classification. We also noticed that in most cases, the proposed schemes scored its best classification rates while using the 5^{th} level of 2D DT-CWT decomposition with a feature vector of 80.

The usage of the BPSO-based fusion (Schemes I, II and V) has resulted in significant performance improvement, while the usage of PCA on the fused feature vector before or after applying the BPSO (Schemes III and IV) has degraded the accuracy rate. The best classification performance allows reducing the original feature space by 97% and hence it also reduces the computation time as compared with conventional methods. This demonstrates that the BPSO based methods allow the same level of performance to be kept while reducing considerably the computation load.

It can be noted that the combination of iris and online signature features has been useful in improving the performance for all the classifiers except for the case from Naïve Bayes classifier in some cases. The performances of the fused features using reduced feature subset indicated that the Naïve Bayesian classifier is very simple and useful, yet it is highly sensitive to feature selection, therefore feature selection is significant.

We have noticed that the accomplished performance of the k-NN classifier suggested that it may be ideal in some applications as it is essentially simple and does not require prior training experience.

One of the important conclusions from Table 5.14 is that the proposed fusion schemes have improved the classification performance rates in terms of accuracy rate and size of feature vectors. The results clearly show that we get reasonable results from the fusion of online signature and iris at the feature level compared with the unimodal systems.

5.6 Summary

In this chapter, we have tackled the problem of feature level fusion in the context of multimodal biometrics. Our concern was to compare different fusion schemes and to provide a clear analysis of their comparative advantages in terms of performance and complexity. With this objective, we considered two independent modalities (iris and online signature) that are represented with different feature extraction techniques. The comparison and combination of proposed features fusion schemes is evaluated on the diverse classification schemes; Naïve Bayes, k-NN and SVM.

The chapter has proposed and investigated the usefulness of Binary Particle Swarm Optimization in a multimodal biometric scenario. The experimental investigations have been shown that we can obtain a considerable improvement in terms of identification performance when Appling the BPSO feature selection scheme to the fused unimodal systems features before performing classification. The implementation of a BPSO algorithm reduced the number of features by a factor of roughly 97% while keeping the same level of performance. Therefore, this approach offers new perspectives for multimodal biometric implementation for biometric traits which are efficiently represented in a high dimension feature space.

Overall, comparing the results with the iris and online signature baselines, it is observed that the feature-level fusion leads to improve the authentication accuracy.

Chapter 6 Hybrid Fusion: Combining Feature and Decision Levels

Decision-level fusion is the most abstract level and consolidates multiple accept/reject decisions from multiple biometric traits to find out the final decision or authentication result. Decision level fusion is the highest level combination possible. This level of fusion takes advantage of the tailored processing performed by each biometric trait. It requires the minimum amount of interaction with user. This chapter studies the performance of decision-level fusion and proposes a new multimodal biometric system based on a hybrid-level-fusion between feature and decision levels in an attempt to improve the final authentication performance.

As discussed in the former chapters, each modality has its strengths and limitations. One approach to improving biometric identification accuracy is to use multiple modalities. Achieving good classification results at the decision level, involves the selection of multiple classifiers and fusion rules that minimize the classification error.

Multiple classifiers systems have been applied to a large number of fields and application domain for decision fusion. Classifier combination is a popular technique in the domain of pattern recognition to improve classification accuracy. In literature, it has been shown that combining classifiers is often practical and effective solution for difficult pattern recognition tasks [65].

The classifier combination approach can be found with different names in literature such as decision combination [66], mixture of experts [79], classifier ensembles [61], classifier fusion [99] consensus aggregation [12], dynamic classifier selection [54], hybrid methods [30] and so on. The difference between these approaches stems mainly from the dependencies between individual classifiers, the selection of classifier outputs, architecture and aggregation strategy. The main benefit of classifier combination is that the performance of classifiers combined is significantly higher than the best obtainable from the individual. In this thesis we shall use the term combing classifiers in its widest meaning, in order to include the whole range of ensemble techniques. This variety of terms and specifications reflects the remarkable effort of the researchers dedicated to this promising discipline.

In this chapter we will consider the combination between classifiers to achieve better detection results through the concept of decision fusion. We study the effect of using a combination of classifiers trained on the different feature sets, over the overall accuracy results.

6.1 Decision Level Fusion

Decision level fusion, also known as fusion at the abstract level [145], considers only classification information of single matchers. Hence, it is possible to apply, as no assumptions about matchers or distributions could be made. This is an advantage as it makes implementation easier. Fusion in a multimodal biometric system is carried out at the decision level when only accept or reject decisions by the individual biometric matchers are available. Figure 6.1 shows the general scheme for decision level fusion. Performing decision fusion therefore means finding the discrete class labels. Although, earlier combination achieves better result than decision level fusion and thus can be more effective [51,154,180]. Nevertheless, this is not always true, as Kumar et al. [98] showed that fusion at decision level outperformed fusion at feature level for multimodal system based on fusion of hand geometry and palmprint.

In this chapter we study the decision fusion in the context of fully automatic iris and online authentication. Firstly decision fusion is used combine the outputs of several iris and online signature authentication algorithms. This type of fusion has been recently studied for different biometric modalities [51,180].

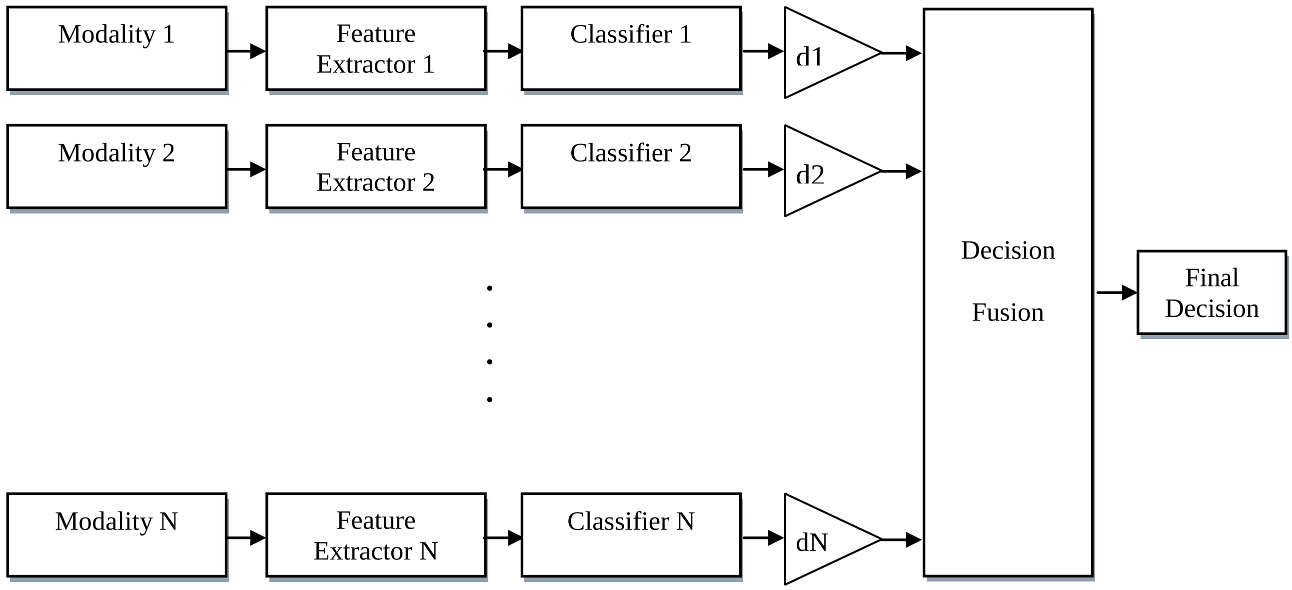


Figure 6.1 Block diagram for decision level fusion

Fusion at decision score level is challenging and less studied in literature, on the basis that decisions have less information content compared to earlier levels of fusion. The majority of the techniques proposed for decision level fusion include majority voting, Bayesian decision fusion, the Dempster-Shafer theory of evidence, "AND" and "OR" rules and weighted majority voting. A brief description of such techniques is presented below.

- **"AND" and "OR" Rules**

Using the "AND" and "OR" rules is the simplest means of combining decisions output by the different matchers. The "AND" rule issues a "match" decision only when all the biometric matchers agree that the claimed identity sample matches with the stored template. In case of "OR" rule, the output is a "match" decision on condition that at least one matcher issues a match decision. When applying the "AND" rule, the FAR is expected to extremely drop compared with the FAR of the individual matchers, whilst the FRR is expected to rise greater than the FRR of the individual matchers. Likewise, the "OR" rule leads to significantly higher FAR and lower FRR than the individual matchers. Thus, it may actually degrade the overall performance of the multimodal biometric system [35].

- **Majority Voting**

This is the most widespread and intuitive approach for decision level fusion where the input biometric sample is assigned to that identity on which the majority of the matchers agree on that identity. Majority voting is based on the assumption that all the matchers perform equally well. This does not require either a priori knowledge about the matchers or any additional training to come up with the final decision. Kuncheva et al. [100] introduced a theoretical analysis of the majority voting fusion scheme by establishing limits on the accuracy of the majority vote rule based on the number of matchers, the individual accuracy of each matcher and the pair wise dependence between the matchers.

- **Weighted Majority Voting**

This technique is usually applied when the recognition accuracy of different matchers are not identical. Therefore, it is reasonable to assign different weights to the decision of different classifiers. Bearing in mind that higher weights are assigned to the decisions made by the more accurate classifiers. In this case the recognition procedure is similar to the majority voting approach, except that the weights of individual classifiers are also considered [154].

- **Bayesian Decision Fusion**

This fusion scheme depends on transforming the discrete decision labels output into continuous probability values. Using Bayes rule, the posterior probability of class w_k $P(w_k|C)$ can be rewritten as

$$P(w_k|c) = \frac{P(c|w_k)P(w_k)}{P(x)} \qquad 6.1$$

Where $P(w_k)$ and $P(x)$ are the a priori probabilities of class i and x, respectively. And $P(c|w_k)$ is the conditional probability of x given c. we will shed more light on this technique in the next section.

- **Dempster-Shafer Theory of Evidence**

Dempster–Shafer evidence is a mathematical theory was developed as an attempt to overcome the limitation of conventional probability theory by handling uncertain, imprecise and incomplete information [36,164]. Dempster–Shafer theory, also known as the theory of belief functions, is often viewed as a generalisation of Baysian probability theory and it is more flexible than Baysian when knowledge is incomplete [164]. Major advantage of this theory is the ability to easily represent evidence at different levels of abstraction and the possibility to combine evidence from different sources. The idea in Dempster-Shafer theory is to build beliefs about the true state of a class from smaller and distinct pieces of evidence.

6.2 The Suggested Decision Level Scenarios

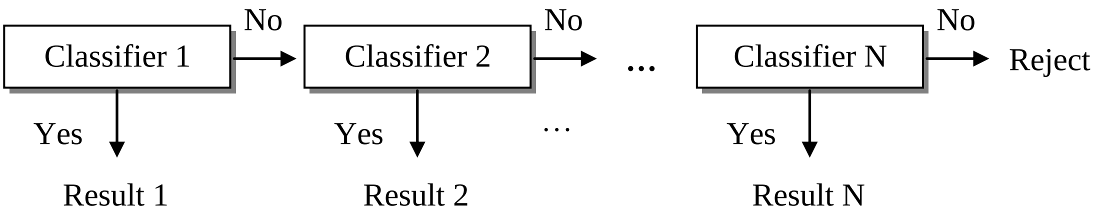
The purpose of this chapter is to investigate the usefulness of decision fusion in the context of fully automatic iris and online authentication.

6.2.1 Architecture of the Individual Classifiers

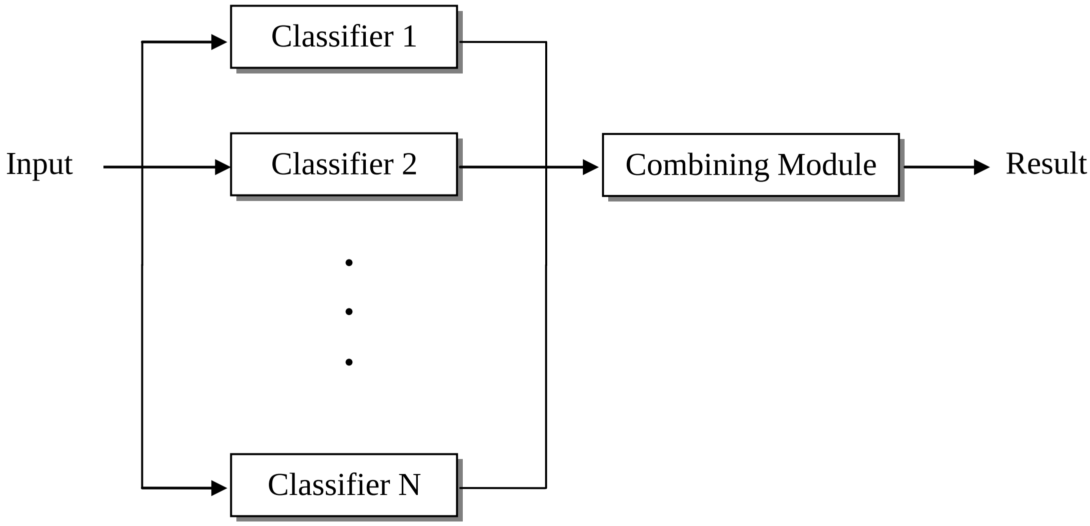
Choosing base classifiers is also very important task in combing classifiers. The composition of the single classifiers will affect its performance [39]; yet, the number of combinations of single classifiers and parameters is almost infinite and thus a thorough evaluation of this experimental factor is outside the scope of this thesis. Therefore, the classifiers that are used are the same those used throughout this thesis which represents a broad number of machine learning approaches. Except we have employed the SVM with the radial basis function (RBF) as the basic kernel function as it seemed to offer the best results in the previous chapters.

6.2.2 Architecture of the multiple classifier system

There are two main approaches being applied in building multiple classifiers: serial and parallel strategies [67]. Serial approach, invoke the classifiers in a cascade order, where some of the classifier may be used only if the first classifier failed to satisfy an acceptable result. On the contrary, in the parallel approach all classifiers are invoked independently with the same input data, and afterwards their decisions are combined. Thus, in serial architecture, the order of classifiers arrangement is critical for the classification performance, whereas in parallel approach, system performance depends mainly on the combination procedure. Moreover, in serial architecture, most of legitimate users can be accepted by using the first biometric in the processing chain, while all available biometrics should be required to the unacceptable users from the first matcher. The method described in this chapter follows the parallel approach. The structures of serial and parallel approaches are shown in Figure 6.2.



(a)



(b)

Figure 6.2 The block diagram of (a) serial and (b) parallel classifier combinations

6.2.3 Combination method

A numerous of possible schemes have been proposed in the literature to combine individual classifiers [75,90,97,98]. Regarding biometrics, it has been shown despite their simplicity, simple combination schemes, have resulted in high recognition rates than trainable fusion rules [91].

The work in this chapter continues in the general direction of combining classifiers based on different feature sets developed in Kittler et al. [90]. In their findings, they state that the sum rule and its elementary fixed combination schemes on measurement level (such as: max rule, majority vote rule, and median rule) consistently outperform other classifier combination schemes. They showed that these elementary combination schemes can be seen as compound classification, where all classifiers are used to make a decision.

In [89], Kittler extends his work analytically to proof that sum fusion strategy outperforms the majority vote when all classifiers are of equal strength and estimation errors are conditionally independent and identically distributed. We briefly introduce the framework in this section.

Suppose N individual classifiers c_n (n =1,....,N) are selected through the classifier-selection step. Each classifier assigns one input sample (represented as $x_k=(x_1,x_2,...,x_N)$) to one of the possible a classes L_k ($L_k = w_1,..., w_m$). Then, according to Bayesian theory, the classifier c_n gives every output a measurement which is represented as a posterior probability vector, $P_n = [p(w_1|w_n), \ldots , p(w_m|w_n)]^t$ where $p(w_i|w_n)$ denotes the probability that the classifier considers that x was labelled with w_i.

The pattern z should be assigned to class w_j provided that a posterior probability is maximum, i.e.

$$assign\ z \longrightarrow w_j$$

$$p(w_j|x_1,...,x_N) = \max_k P(w_k|x_1,...,x_N) \qquad 6.2$$

From the Bayes theorem, the posteriori probability can be rewritten as

$$P(w_k|x_1,...,x_N) = \frac{p(x_1,...,x_N|w_k)P(w_k)}{P(x_1,...,x_N)} \qquad 6.3$$

where $p(x_1,...,x_N|w_k)$ is the conditional joint probability density function for measurements on class w_k and $p(x_1,...,x_N)$ is the unconditional measurement joint probability function.

6.2.4 Classifier Algebraic Combination Strategies

- **Sum rule**

The sum rule can be derived if we assume that the a posteriori probabilities computed from the classifiers do not differ greatly from the a priori probabilities. Then the a posteriori probabilities can be expressed as

$$P(w_k|x_i) = P(w_k)(1+\delta_{ki}) \tag{6.4}$$

where $\delta_{ki} << 1$. Substituting the a posteriori probability in equation (6.8) into the posteriori probabilities in equation (6.9),

$$P^{-(N-1)}(w_k)\prod_{i=1}^{N} P(w_k|x_i) = P(w_k)\prod_{i=1}^{N}(1+\delta_{ki}) \tag{6.5}$$

By expanding the product and neglecting terms of the second and higher order in δ_{ki},

$$P^{-(N-1)}(w_k)\prod_{i=1}^{N} P(w_k|x_i) = P(w_k) + P(w_k)\sum_{i=1}^{N}\delta_{ki} \tag{6.6}$$

Then using (6.8) to eliminate δ_{ki}, we obtain the sum decision rule as

$assign\ z \longrightarrow w_j$ $\quad$ if

$$(1-N)P(w_j) + \sum_{i=1}^{R} P(w_j|x_i) = \max_{k=1}^{m}[(1-N)P(w_k) + \sum_{i=1}^{N} P(w_k|x_i)] \tag{6.7}$$

- **Max rule**

Starting from (6.7) and approximating the sum by the maximum of the posterior probabilities, we obtain the max rule as

assign $z \longrightarrow w_j$ $\quad$ if

$$(1-N)P(w_j) + N\max_{k=1}^{N} P(w_j|x_i) = \max_{k=1}^{m}\left[(1-N)P(w_k) + R\max_{i=1}^{R} P(w_k|x_i)\ \right] \tag{6.8}$$

which, with the assumption of equal a priori probabilities, becomes

$assign\ z \longrightarrow w_j$ $\quad$ if

$$\max_{k=1}^{N} P(w_j|x_i) = \max_{k=1}^{m}\max_{i=1}^{R} P(w_i|x_i) \tag{6.9}$$

- **Minimum rule**

Starting from (6.7) we obtain a minimum decision rule

assign $z \longrightarrow w_j$ *if*

$$P^{-(N-1)}(w_j)\min_{i=1}^{N} P(w_k|x_i) = \max_{k=1}^{m} P^{-(N-1)}(w_k)\min_{i=1}^{N} P(w_k|x_i) \qquad 6.10$$

with the assumption of equal a priori probabilities, a minimum decision rule becomes

assign $z \longrightarrow w_j$ if

$$\min_{i=1}^{N} P(w_k|x_i) = \max_{k=1}^{m} \min_{i=1}^{N} P(w_k|x_i) \qquad 6.11$$

- **Majority voting rule**

Starting from (6.7) under the assumption of equal priors and if the a posteriori probabilities are hardened to produce a binary valued function,

assign $z \longrightarrow w_j$ *if*

$$j = \max_{k=1}^{m} \sum_{i=1}^{R} \Delta_i \qquad 6.12$$

where

$$\Delta_i = \begin{cases} 1 & L_k = w_i \\ 0 & L_k \neq w_i \end{cases}$$

And at the end the class with the largest number of votes is selected.

6.3 Decision-Level Fusion System

The basic idea here is to fuse the decisions of the individual iris and online signature biometrics. Each biometric decision was evaluated by the three classifiers: SVM, Naïve Bayes and k-NN. In a multi-classifier decision fusion context, each classifier has a decision, the decisions from multiple classifiers are then fused in order to generate the final decision. The input of each classifier is a vector composed of values of the selected features and the output is a class label of the sample. Let $F_{signature}$ and F_{iris} denote the feature vectors of the online signature and iris respectively. The combined decision can be obtained into two steps, first by training each classifier independently on the same feature set and thus, obtaining an individual decision using the well-known fixed rules.

To combine the final decision of the individual classifiers in order to find $D_{signature}$ and D_{iris} the estimated class w_i, given input $F_{signature}$ and F_{iris} is given by

$$D_{iris} = \Xi(D_{iris\_SVM}, D_{iris\_k-NN}, D_{iris\_NaiveBayes}) \quad 6.13$$

$$D_{signature} = \Xi(D_{signature\_SVM}, D_{signature\_k-NN}, D_{signature\_NaiveBayes}) \quad 6.14$$

where Ξ is the selected combining rule (i.e. maximum, sum, majority or minimum rule) evaluated in this chapter.

One of the weaknesses of fixed rules is the fusion of the decisions of the individual classifiers is based on assumption that the classifiers are independent. This assumption may be quite suited, especially for the iris and online signature based features. Therefore AND rule can be better alternative for consolidating single decisions (Figure 6.3) as the AND rule is estimated to perform better on the assumption of independent data representation [35,51,146]. Therefore, we decided to fuse the decisions by the AND rule to obtain the final decision.

To combine the final decision of the multi-classifiers in order to obtain the $D_{FinalClass}$ the estimated class w_i, given $D_{signature}$ and D_{iris} as inputs, is as follows.

$$D_{FinalClass} = And(D_{iris}, D_{signature}) \quad 6.15$$

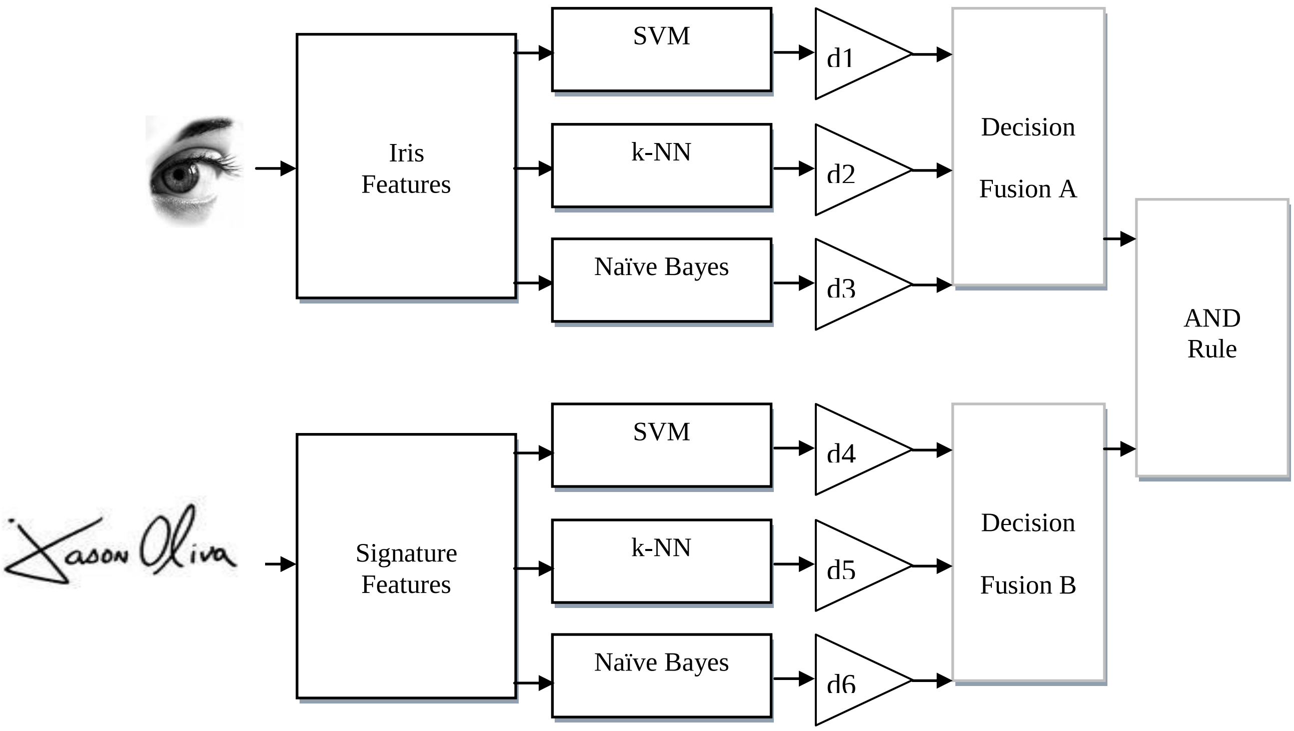


Figure 6.3 Schematic for proposed multimodal decision-level fusion scheme

6.4 Hybrid Fusion System

In an attempt to improve the final authentication performance, we further propose a hybrid fusion technique, which combines the feature-level and decision-level fusions, taking advantage of both fusion modes. The motivation behind the suggested hybrid fusion is twofold. Firstly, we demonstrate that the decision fusion framework can be integrated easily. Secondly, by hybrid fusion we expect to take advantage of the feature-level and decision-level fusion, and eventually achieve more reliable and robust biometric system.

We summarize the hybrid fusion method as follows: the proposed scheme starts first with performing the PCA to reduce the size of S_{Iris} $S_{Signature}$ independently before vertically fusing both features into one feature vector $X_{FusedFeatures}$. Followed by reducing the dimension of the fused features $\overline{X}_{FusedFeatures}$ using the BPSO. Afterwards, each classifier vector will be trained and tested independently with the fused feature sets and the output is the class label of the sample. To combine the final decision of the individual classifiers in order to obtain the $D_{FinalClass}$ the estimated class w_i, given input $\overline{X}_{FusedFeatures}$ is given by

$$D_{FinalClass} = \Xi(D_{FusedFeatures\_SVM}, D_{iFusedFeatures\_k-NN}, D_{FusedFeatures\_NaiveBayes}) \qquad 6.16$$

Where Ξ is the selected combining rule (i.e. maximum, sum, majority or minimum rule) evaluated in this chapter, as shown in Figure 6.4.

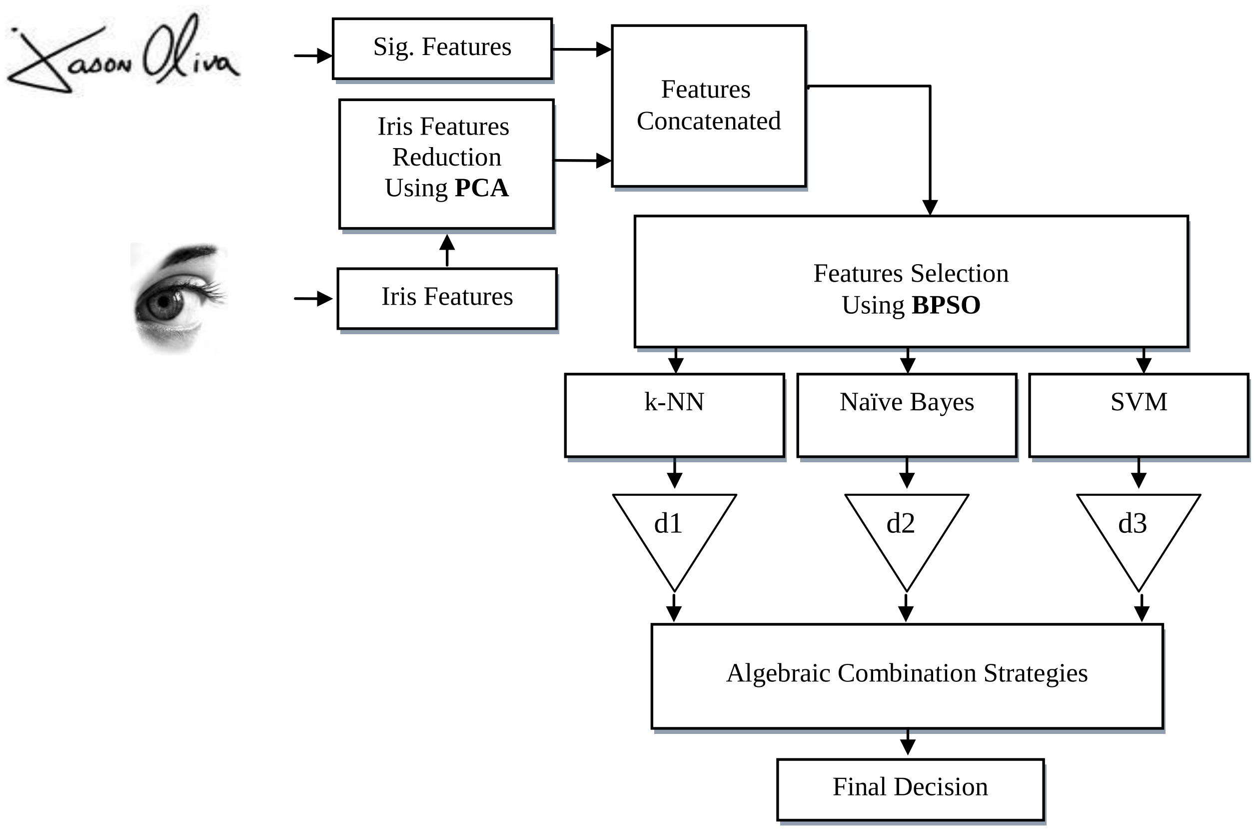


Figure 6.4 Schematic for proposed hybrid multimodal fusion scheme

6.5 Experimental Results

6.5.1 Decision-level Fusion Scheme

This section describes the experimental setup, including the assessment protocol that we have built in order to evaluate the proposed decision and hybrid level fusion schemes. For all of the experiments in this chapter, the same underlying conditions applied in the previous chapters have been carried out in this chapter. In the first decision fusion scheme, the individual classifiers were trained and tested first with the 31 global functions of the online signatures along with the 2D-DTCWT iris features. Then the individual decisions from the three classifiers combined with maximum, sum, majority and minimum rule. Finally the resulted two decisions were combined with the AND rule. Here is a description of the steps involved in the first scheme

Step 1: Extract and normalize the iris image
Step 2: Apply the 2DT-CWT on the extracted iris images to obtain S_{Iris} wavelet coefficients,
Step 3: Carry out the iris classification using the three classifiers independently:
- *SVM (RBF Kernel)*
- *k-NN*
- *Naïve Bayes.*

Step 4 : Combine the three decisions in step 3 with the four algebraic rules (maximum, sum, majority ,minimum) at each time
Step 5: Extract the 31 global features from the dynamic signatures $S_{Signature}$.
Step 6: Carry out the online signature classification using the three classifiers independently:
- *SVM (RBF Kernel)*
- *k-NN*
- *Naïve Bayes.*

Step 7 : Combine the two decisions in step 6 with the four algebraic rules (maximum, sum, majority ,minimum) at each time
Step 8 : Combine the two decisions from steps 4 and 7 with the AND rule.

Table 6.1 presents the summary of experimental results of the first scheme, i.e., percentage recognition rate using different combination schemes considered in this chapter. As previously stated, we plan to design a simple iris-signature multi biometrics system based on decision level fusion.

Table 6.1 Performance rates from the proposed decision level fusion, Scheme I

Biometrics			Combination Strategy			
Online signature	Iris	Dim.	Sum	Majority	Minimum	Maximum
31	1040	1071	98.97	98.93	98.94	98.39
31	280	311	98.96	98.96	98.91	98.63
31	80	111	98.45	98.46	98.49	98.36
31	24	55	97.38	97.41	97.41	97.39

The best performance is achieved by the Sum rule with the combination of the online signatures with the iris of the 3rd and 4th level. The combination of the 6th level of 2DT-CWT with the online signature features has been worst and this may be due to the small number of features that make the classification of data difficult. The performance of Sum rule is better than Minimum and Maximum rules but quite similar to that from Majority rule. One of the important conclusions from the Table is that the proposed is that the proposed framework has effectively reduced the number of features by 89.63% while improving or maintaining similar performance in most cases.

This surely suggests that only a small subset of iris features is necessary in practice for building an accurate model for authentication. The performed experiments over the decision level fusion using fixed combination rules suggests that the performance from the Sum, Majority and Minimum rules have been effective in improving the performance.

The second decision fusion scheme is quite similar to the first scheme except that the online signature individual classifiers were trained and tested first with the reducted set of the global functions. Here is a description of the steps involved in the second scheme:

Step 1: Extract and normalize the iris image
Step 2: Apply the 2DT-CWT on the extracted iris images to obtain S_{Iris} wavelet coefficients,
Step 3: Carry out the iris classification using the three classifiers independently:

- *SVM (RBF Kernel function)*
- *k-NN*
- *Naïve Bayes.*

Step 4 : Combine the three decisions in step 3 with the four algebraic rules (maximum, sum, majority ,minimum) at each time
Step 5: Extract the 31 global features from the dynamic signatures.

Step 6: Reduct the dynamic signatures $S_{Signature}$ using the Rough set.
Step 7: Carry out the online signature classification using the three classifiers independently:

- *SVM (RBF Kernel function)*
- *k-NN*
- *Naïve Bayes.*

Step 8 : Combine the two decisions in step 7 with the four algebraic rules (maximum, sum, majority ,minimum) at each time
Step 9 : Combine the two decisions from steps 4 and 8 with the AND rule.

Table 6.2 Performance rates from the proposed decision level fusion, Scheme II

Biometrics			Combination Strategy			
Online signature	Iris	Dim.	Sum	Majority	Minimum	Maximum
9	1040	1049	97.41	98.77	98.88	98.30
9	280	289	98.81	98.80	98.79	98.49
9	80	89	98.38	98.37	98.41	98.30
9	24	33	97.30	97.30	97.29	97.29

The best performance is achieved by the Sum rule with the most combinations of the 9 features of the online signatures with the iris features. The combination of the 6th level of 2DT-CWT with the online signature features has been worst and this may be due to the small number of features (33 in this case) that make the classification of data difficult.

The performance of Sum, Minimum and Maximum rules rule is better than Majority rule. One of the notable remarks extracted from Table 6.1 is that the proposed framework has effectively reduced the number of features by 91.51% while maintaining a similar performance to the best cases.

The performed experiments over the decision level fusion using fixed combination rules suggests that the performance from the Sum, Majority and Minimum rules have been effective in improving the performance.

One major advantage of this fusion level scheme is that fusing multimodal biometric features at decision level with AND rule is simple and feasible. As the two modal features are extracted from different parts of the body, thus, the two levels of features are not correlated and the classifiers in this case are independent which can ensure the AND rule to enhance the authentication performance theoretically [35,154].

6.5.2 Hybrid Fusion Scheme

The proposed hybrid fusion technique aims to improve the final authentication performance taking advantage of the feature and decision-levels. The steps involved in the hybrid fusion technique include the following steps:

Step 1: Apply the 2DT-CWT on the extracted iris images to obtain S_{Iris} wavelet coefficients.
Step 2: Extract the global features from the dynamic signatures $S_{Signature}$.
Step 3: Apply PCA to the iris feature vector S_{Iris} to obtain S_{IrisR}.
Step 4: Vertically concatenate S_{IrisR} and $S_{Signature}$ to obtain the fused features vector $X_{FusedFeatures}$.
Step 5: Normalize the fused feature using min-max normalization to obtain $\overline{X}_{FusedFeatures}$.
Step 6: Randomly initialize the PSO particles with binary values (0 and 1).
Step 7: Carry out the feature selection by considering the value of the bit in the particle. More precisely, if bit value is 1, select the corresponding feature from $\overline{X}_{FusedFeatures}$.This way we construct a new feature vector $\overline{X}_{NewFusedFeatures}$.
Step 8: Carry out the classification using the three classifiers independently with the new feature vector $\overline{X}_{NewFusedFeatures}$:

- *SVM (RBF Kernel function)*
- *k-NN*
- *Naïve Bayes.*

Step 9 : Combine the three decisions in step 8 with one of the four algebraic rules (maximum, sum, majority ,minimum) one at each time.

In the first set of experiments, the decision fusion scheme is applied where the single classifiers were trained and tested first with the 31 global functions of the online signatures and the 2D-DTCWT iris features. Whilst, the second set of experiments the decision fusion scheme is applied by the single classifiers were trained and tested first with the nine global functions of the online signatures. Tables 6.3 and 6.4 present the results of the hybrid fusion approach of the first and the second set of experiments respectively.

The best performance, in both set of experiments, is achieved by the Sum rule with the combination between the online signatures with the 6^{th} level of 2DT-CWT iris features. In most cases, the performance of Sum and Majority rules is better than Minimum and Maximum rule.

The performed experiments over the decision level fusion using fixed combination rules suggests that the performance from the Sum and Majority rules have been effective in improving the overall performance. It can be observed that the suggested hybrid fusion scheme performed better than the decision-level fusion. It is interesting to note that this strategy has been able to achieve an accuracy of 99.73% with 22 features only.

Table 6.3 Performance rates from the hybrid fusion scheme I

Biometrics Online signature	Iris	Dim.	Combination Strategy Sum	Majority	Minimum	Maximum
31	1040	86	97.75	97.75	97.35	96.42
31	280	76	97.61	97.75	97.88	96.82
31	80	41	99.33	99.47	99.33	98.41
31	24	22	99.73	99.33	99.20	99.20

Table 6.4 Performance rates from the hybrid fusion scheme II

Biometrics Online signature	Iris	Dim.	Combination Strategy Sum	Majority	Minimum	Maximum
9	1040	254	78.30	75.13	70.23	78.04
9	280	76	95.10	94.97	94.57	94.17
9	80	22	87.56	87.43	86.50	86.50
9	24	16	97.22	97.08	95.76	96.03

Table 6.5 Proposed schemes recognition rates (%)

Method	Classifier/Scheme	Dimension	Recognition Rate (%)
Iris *alone*	SVM-RBF	1040	92.86
	SVM- Polynomial	1040	92.46
	SVM-Linear	1040	92.32
	k-NN	1040	80.82
	Naïve Bayes	280	77.11
Online signature *alone*	k-NN	31	98.33
	k-NN	9	95.41
	Naïve Bayes	31	97.10
	Naïve Bayes	9	96.30
	SVM-RBF	31	98.54
	SVM- Polynomial	31	98.54
	SVM-Linear	31	98.94
	SVM-RBF	9	95.76
	SVM- Polynomial	9	95.50
	SVM-Linear	9	95.23
Iris *and* Online signature			
Feature Fusion	Scheme V_a	18	98.94
	Scheme II_b	72	96.29
Decision Fusion	Sum Rule	1049	97.41
	Sum Rule	289	98.81
Hybrid Fusion	Sum Rule	22	99.73
	Sum Rule	16	97.22

As discussed in [127,154], it is usually believed that applying the combination strategy at an early stage of the integration stage can guarantee better performance results. As the feature-level contains more information about the unknown biometric patterns, thus this level is expected to provide better performance than the decision-level combination. This agrees with the results we obtained where the feature level fusion outperformed the decision level fusion.

A number of important outcomes of the experimental analysis can be observed by considering the results in all the tables shown above. From these results, it is clearly seen that the best performance is noted for hybrid fusion scheme I_a was a GAR of 99.73% with the Sum rule with a feature vector of size 22 which outperforms the feature level schemes in terms of accuracy rate and size of feature vector. The best classification performance allows reducing the original feature space by 98% and hence it also reduces the computation time as compared with conventional methods. This demonstrates that the BPSO based methods allow the same level of performance to be kept while reducing considerably the computation load.

The experimental results showed that the Sum rule achieved the best performance which synchronizes with the results obtained by [51,89]. More importantly, the hybrid fusion scheme with the 31 online signature features showed a better performance as compared with the hybrid fusion scheme with the reducted set of online signature features. This clearly indicates that the number of online signature features plays a significant role in classification. We also noticed that in most cases, the proposed schemes scored its best classification rates while using the 6th level of 2D DT-CWT decomposition with a feature vector of 24.

Table 6.6 Performance of some multimodal systems

Authors	Biometric traits	Recognition rate
Yao et al. [189]	Face and palmprint	90.73%
Kumar et al. [97]	palmprint and hand geometry	98.59%
Jain et al. [171]	Face, fingerprint and hand-geometry	98.6%
Bergamini et al. [14]	Face and fingerprint	99.80%
Nandakumar et al. [154]	Fingerprint and Iris	94.8%
Proposed-*feature-level* fusion	Iris and online signature	98.94
Proposed-*decision-level* fusion		98.81
Proposed-*hybrid* fusion		99.73

Table 6.6 summarizes the performance of some of the reported multimodal systems that have been examined by a number of researchers. We can notice that the best-reported accuracy rate is similar to suggested hybrid fusion technique. One of the important conclusions from Tables 6.3-6.6 is that the proposed multimodal biometric authentication schemes achieved promising results and –at the same time- improved the classification performance rates in terms of accuracy rate and size of feature vectors. The results clearly show that we got reasonable results from the fusion of online signature and iris compared with the unimodal systems and feature and decision levels.

6.6 Summary

The objective of this work was to investigate the integration of online signature and iris features, and to achieve a better performance that may not be achievable with single biometric alone. The experimental investigations have been concerned with the fusion of online signature and iris biometrics in the decision and hybrid fusion modes. The basic idea was to fuse and evaluate the decisions of the SVM, Naïve Bayes and k-NN classifiers using fixed rules: Maximum, Sum, Majority and Minimum rules. The individual decisions from the two modalities were further combined with the AND logic rule to obtain the final decision. The AND logic was applied to ensure a satisfactory level of security, since a positive authentication is only accomplished in case if only all the fusion levels approaches produce positive authentication [35,80].

In an attempt to improve the final authentication performance, we further proposed a hybrid fusion technique, which combines the feature-level and decision-level fusions, taking advantage of both fusion modes. The motivation behind the suggested hybrid fusion is twofold. Firstly, we demonstrate that the decision fusion framework can be integrated easily within the authentication procedure. Secondly, we expect to take advantage of the feature-level and decision-level fusion, and

eventually achieve more reliable and robust biometric system. Based on the experimental investigations, it has been shown that the hybrid approach offers considerable improvements to the accuracy of multimodal biometrics.
The experimental results presented on the chimeric database suggest that the proposed feature-decision-level combination approach can be effectively employed to achieve the performance improvement from the feature or decision level alone.

We remark that our suggested biometric system was able to achieve good accuracy recognition results. We have been able to reduce the feature vector yet, while keeping enough discriminatory power to be used as a possible biometric in recognition applications.

One of the main advantages of the suggested framework is that the individual classifiers can be trained separately; thus extending a multimodal system to incorporate new modalities is effortless. In addition, the process of collecting and training data separately is more practical more than collecting and training the whole dataset.

Chapter 7

Conclusion and Future Work

Even though further work remains to be done, our results to date indicate that the combination of online signature and iris features represents a promising addition to the biometrics-based personal authentication systems. Our experimental results demonstrate that while majority of iris and online signature characteristics are useful in predicting the person's identity, only a small subset of these features are required in practice for building an accurate model for authentication.

7.1 Research Summary

The work in this thesis can be summed up as follows:

At the start we introduced the topic of biometric and main characteristics and challenges of Biometrics. Later we investigated the key issues in multimodal biometric systems along with the different architectures for information integration, and review the previous investigations in multimodal biometrics. We have observed that, only limited work is reported on feature level fusion of multimodal biometric system. Furthermore, we also noticed there is no reported research work that combines iris and online signature (Chapter 2).

The first step towards our goal was to build an online signature authenticating system using global features. We described our work on building an online signature database and performing statistical analysis of online signature signals. An online signature authentication algorithm based on comparing the performance of three feature selection algorithms was constructed for the selected feature set (Chapter 3).

The next step was to develop a novel iris segmentation approach based on minimizing the effect of the eyelids and eyelashes by trimming the iris area above the upper and the area below the lower boundaries of the pupil. To increase the recognition accuracy we extracted the 2D dual-tree complex wavelet transform from the iris images. The proposed features was evaluated by a diverse classification schemes namely; Naïve Bayes, k-NN and SVM. The approach was evaluated on a benchmark iris dataset (Chapter 4).

Afterwards, we proposed and investigated the usefulness of Binary Particle Swarm Optimization in a range of feature-level fusion scenarios between iris and online signature. The experimental investigations have been shown that we can obtain a considerable improvement in terms of identification performance when Appling the BPSO feature selection scheme to the fused unimodal systems features before performing classification. In general, comparing the results with the iris and online signature baselines, it is observed that the feature-level fusion leads to the improvement of the authentication accuracy (Chapter 5).

Next, an experimental investigation is conducted on the fusion of online signature and iris biometrics at the decision fusion mode. The basic idea was to fuse and evaluate the decisions of the SVM, Naïve Bayes and k-NN classifiers using fixed rules: Maximum, Sum, Majority and Minimum rules. The individual decisions from the two modalities were further combined with the AND logic rule to obtain the final decision. Finally, in an attempt to improve the final authentication performance, we proposed a hybrid fusion technique, which combines the feature-level and decision-level fusions, taking advantage of both fusion modes. We remarked that our suggested biometric system was able to achieve good accuracy recognition results. We have been able to deduce the feature vector yet, while keeping enough discriminatory power to be used as a possible biometric in recognition applications (Chapter 6).

7.2 Contribution to Knowledge

The essential objective of current research work is to examine whether the performance of a biometric system can be improved by integrating complementary information which comes primarily from two different and independent modalities. Therefore, this thesis makes the following main original contributions.

A Novel Online Signature Authentication Approach

A novel online signature identification scheme based on global features and Rough set is proposed. The information was extracted as time functions of various dynamic properties of the signatures. Rough set approach has resulted in a reduced set of nine features that were found to capture the essential characteristics required for signature identification. The reported results demonstrate the suitability and effectiveness of the Rough set approach in the application of online signature identification.

Iris Authentication Technique using 2D Dual-Tree Complex Wavelet Transform and Support Vector Machine

Iris patterns are believed to be unique due to the complexity of the underlying the environmental and genetic processes that influence the generation of iris pattern. Segmenting iris area is a challenging task since the iris region can be occluded by eyelids or eyelashes. In this thesis we proposed new iris segmentation approach based on minimizing the effect of the eyelids and eyelashes. The dual-tree complex wavelet transform was extracted from the iris images and used to increase the recognition accuracy. The proposed innovative technique proofed to be computationally effective as well as reliable in term of recognition compared with other techniques.

Hybrid Fusion: Combining Feature and Decision-Levels

The experimental investigations have been concerned with the fusion of online signature and iris biometrics in the decision and hybrid fusion modes. The individual decisions from the two modalities were further combined with the AND logic rule to obtain the final decision.

In an attempt to improve the final authentication performance, we further proposed a hybrid fusion technique, which combined the feature-level and decision-level fusions, taking advantage of both fusion modes. Based on the experimental investigations, it has been shown that the hybrid approach offers considerable improvements to the accuracy of multimodal biometrics.

7.3 Success Criteria Revisited

To answer the research questions that we pointed out in Chapter 1, an automated multimodal biometric authentication system have been built and tested throughout the thesis. The objective of this work was to investigate the integration of online signature and iris clues, and to achieve a better performance that may not be achievable with single biometric alone. The experimental investigations, which combined the feature-level and decision-level fusions, have improved the final authentication performance. Therefore, it has been shown that the proposed hybrid approach offers considerable improvements to the accuracy of multimodal biometrics.

7.4 Future Work

So far, the issue of recognising people by using iris and online signature has been thoroughly discussed. Despite its promise, which has been shown in this thesis, in this section we discuss possible directions for future research.

1. To begin with, the authentication results presented in this thesis should be validated using other public multimodal real-user databases. Specifically, it would be necessary to measure the performance of the suggested approaches with a larger dataset, containing more individuals. Unfortunately, as far our knowledge is concerned, there are no public real-user database which combines online signature and iris modalities of the same individuals available that could be suited to evaluate our schemes.

2. In the multimodal biometric literature a lot of attention has been paid to parallel fusion of multiple classifiers. A few of reported works dealt so far with serial architecture. Serial approach, invoke the classifiers in a cascade order, where some of the classifier may be used only if the first classifier failed to satisfy an acceptable result. While, in the parallel approach, all classifiers are invoked independently with the same input data, and afterwards their decisions are combined. It would also be of interest to study the performance of the proposed techniques with the serial fusion of multiple classifiers.

3. The proposed techniques in this thesis can also be applied with other kinds of biometrics. It would be interesting to integrate other behavioural and physiological biometrics such as palm vein and face in conjunction with iris and online signature biometrics to enhance recognition performance.

4. A future way to improve the recognition system introduced in Chapter six could be to develop an alternative space reduction strategy that demonstrates better discriminative properties than BPSO such as Markov Blanket filtering algorithm.

5. As present investigation is only limited to identification accuracy, future research could be to extend for possible verification applications.

Faculty of Technology

De Montfort University

England, United Kingdom
February, 2012

References

1. P. Addison. The Illustrated Wavelet Transform Handbook. Institute of Physics, 2002.

2. F.A. Afsar, M. Arif , U. Farrukh. Wavelet Transform Based Global Features for Online Signature Recognition. Proceeding of IEEE International Multi-topic Conference INMIC, pp. 1-6, 2005.

3. D. W. Aha, D. Kibler, M. K. Albert. Instance based learning algorithms. Machine Learning, Vol.6, pp. 37-66, 1991.

4. W. Almayyan, H.S. Own, H. Zedan. Iris features extraction using dual-tree complex wavelet transform. International Conference of Soft Computing and Pattern Recognition (SoCPaR 2010), pp.18-22, 2010.

5. W. Almayyan, H.S. Own, H. Zedan. Information Fusion in Biometrics: A Case Study in Online Signature. The International Multi-Conference on Complexity, Informatics and Cybernetics: IMCIC,2010.

6. W. Almayyan, H.S. Own, R. Ramadan, H. Zedan. A Multimodal Biometric Fusion Approach based on Binary Particle Optimization. Proceedings of AI-2011 Thirty-first SGAI International Conference on Artificial Intelligence, Cambridge, England, pp. 139- 152, 2011.

7. W. Al-Mayyan, H.S. Own, H. Zedan. Rough set approach to online signature identification. Digital Signal Processing, Vol.21(3), pp.477-485 , 2011.

8. R.J. Anderson. Security Engineering: A Guide to Building Dependable Distributed Systems. New York: Wiley, 2001.

9. J. Barros, J. French, W. Martin. Indexing Multi-Spectral Images for Content-Based Retrieval. University of Virginia Technical Report, CS-94-40, 1994.

10. J. Barros, J. French, W. Martin. System for indexing multi-spectral satellite images for efficient content-based retrieval. Proceedings of the SPIE, Vol. 2420, pp. 228–237, 1995.

11. J. Bazan, H.S. Nguyen, S.H. Nguyen, P. Synak, J. Wróblewski. Rough set algorithms in classification problem. In: L. Polkowski, S. Tsumoto, T.Y. Lin (Eds.), Rough Set Methods and Applications, Physical Verlag, pp. 49–88, 2000.

12. J. Benediktsson, P. Swain. Consensus theoretic classification methods. IEEE Transactions on System and Man Cybernetic, Vol.22 (4), pp.688–704,1992.

13. S. Ben-Yacoub, Y. Abdeljaoued, E. Mayoraz. Fusion of Face and Speech Data for Person Identity Verification. IEEE Transactions on Neural Networks, Vol.10(5), 1999.

14. C. Bergamini, L. Oliveira, A. Koerich, and R. Sabourin. Combining different biometric traits with one-class classification. Signal Processing, Vol.89(11), pp. 2117–2127, 2009.

15. Biometrics History. NSTC. Home page, http://www.biometrics.gov/Documents/BioHistory.pdf.
Last visited 30th January 2011.

16. W.W. Boles , B. Boashash. A Human Identification Technique Using Images of the Iris and Wavelet Transform. IEEE Transactions on Signal Processing, Vol. 46(4), pp.1185-1188, 1998.

17. B. Bonabeau, M. Dorigo, G. Thraulaz. Swarm intelligence: from natural to artificial systems. Oxford University Press, 1999.

18. B.E. Boser, I. Guyon, V. Vapnik. A Training Algorithm for Optimal Margin Classifiers. Proceedings of COLT, pp.144-152,1992.

19. K. Bowyer, K. Hollingsworth, P. Flynn. Image Understanding for Iris Biometrics: A Survey. Computer Vision and Image Understanding, Vol. 110(2), pp.281–307, 2008.

20. J. Bromley, I. Guyon, Y. LeCun, E. Sackinger, R. Shah. Signature Verification using a Siamese Time Delay Neural Network. Advances in Neural Information Processing Systems, Vol. 6, (J. Cowan and G. Tesauro, eds.), 1993.

21. Center for Biometrics and Security Research, CASIA Iris Image Database. Home page, http://www.sinobiometrics.com, last visited 10th January 2012.

22. K. Chang, K.W. Bowyer, B. Victor. Comparison and combination of ear and face images for appearance-based biometrics. IEEE Transactions on Pattern Matching and Machine Intelligence, Vol.25(9), pp.1160–1165, 2003.

23. C. C. Chang, C. J. Lin. LIBSVM: A library for support vector machines. 2001.

24. K. Chang, K. W. Bowyer, S. Sarkar, B. Victor. Comparison and Combination of Ear and Face Images in Appearance-based Biometrics. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.25(9), pp.1160-1165,2003.

25. Y. Chen, S. C. Dass, and A. K. Jain. Fingerprint Quality Indices for Predicting Authentication Performance. Proceedings of the Fifth International Conference on Audio and Video-Based Person Authentication, (AVBPA'05), pp. 160–170, 2005.

26. W. Chen, S. Yuan. A novel personal biometric authentication technique using human iris based on fractal dimension features. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'03), Vol.3 pp. 201-204, 2003.

27. L. Chuang, C. Yang, J. Li. Chaotic maps based on binary particle swarm optimization for feature selection. Applied Soft Computing, Vol.11(1), pp.239-248,2011.

28. R. Clarke. Biometrics and Privacy. 2001.

29. P. Cunningham, S. J. Delany. k-Nearest Neighbour Classifiers. University College Dublin, Technical Report UCD-CSI-2007-4, March 27, 2007.

30. B. Dasarathy. Decision Fusion. IEEE Computer Society Press, Silver Spring, MD, 1994.

31. J. Daugman , C. Downing. Epigenetic randomness, complexity and singularity of human iris patterns. Proceedings of the Royal Society of London - B, Vol.268, pp.1737–1740, 2001.

32. J. Daugman. How iris recognition works. IEEE Transactions on Circuits and Systems for Video Technology, Vol.14(1), pp.21-30, 2004.

33. J. Daugman. Iris recognition. American Scientist, Vol.89(4), pp.326–333, 2001.

34. J. Daugman. High Confidence Visual Recognition of Persons by a Test of Statistical Independence. IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.15(11), pp.1148-1161, 1993.

35. J. Daugman. Combining multiple biometrics. Home page, http://www.cl.cam.ac.uk/users/jgd1000/combine/combine.html , 2000 , last visited 15th November 201.

36. A. Dempster. Upper and lower probabilities induced by multivalued mapping. Ann. Math. Statist. 38, 325–339. 1967.

37. D. Dessimoz, J. Richiardi, C. Champod, A. Drygajlo. Multimodal biometrics for identity documents (MBioID): State-of-the-art (version 2.0). Research Report PFS 341-08.05, 2006.

38. Y.B. Dibike, S. Velickov, D.P. Solomatine, M.B. Abbott. Model induction with support vector machines: introduction and applications. Journal of Computing in Civil Engineering, Vol.15, pp. 208–216, 2001.

39. T.G. Dietterich. Ensemble methods in machine learning. In J. Kittler and F. Roli, editor. Proceedings of the First International Workshop on Multiple Classifier Systems, pp.1-15, 2000.

40. P. Domingos, M. Pazzani. On the optimality of the simple Bayesian classifier under zero–one loss. Proceedings of Machine Learning, Vol.29, pp. 103–130,1997.

41. M. Faundez-Zanuy. Data fusion in biometrics. IEEE Aerospace and Electronic Systems Magazine, Vol. 20, pp.34-38, 2005.

42. M. Faundez-Zanuy. On-Line Signature Recognition Based on VQ-DTW", Pattern Recognition, Vol. 40(3), pp. 981-992, 2007.

43. G. Feng, K. Dong, D. Hu, D. Zhang. When faces are combined with palmprints: a novel biometric fusion strategy. Proceedings of the First International Conference on Biometric Authentication (ICBA), pp. 701–707, 2004.

44. J. Fierrez-Aguilar. Adapted Fusion Schemes for Multimodal Biometric Authentication. PhD thesis, Universidad Politecnica de Madrid, 2006.

45. J. Fierrez-Aguilar, J. Ortega-Garcia, D. Garcia-Romero and J. Gonzalez-Rodriguez. A comparative evaluation of fusion strategies for multimodal biometric verification. Springer LNCS-2688, Fourth International Conference on Audio- and Video-Based Biometric Person Authentication (AVBPA 2003)Guildford, pp.830–837, 2003.

46. J. Fierrez-Aguilar, L. Nanni, J. Lopez-Penalba, J. Ortega-Garcia, D. Maltoni. An online signature verification system based on fusion of local and global information. Proceedings of IAPR International Conference on Audio- and Video-Based Biometric Person Authentication, pp. 523–532, 2005.

47. J. Fierrez-Aguilar, J. Ortega-Garcia, D. Garcia-Romero, J. Gonzalez-Rodriguez. A comparative evaluation of fusion strategies for multimodal biometric verification. Proceedings of the 4th International Conference Audio- Video-Based Biometric Person Authentication, J. Kittler and M. Nixon, Eds., Vol. LNCS 2688, pp. 830–837, 2003.

48. J. Fierrez-Aguilar, J. Ortega-Garcia, J. Gonzalez-Rodriguez, J. Bigun. Discriminative Multimodal Biometric Authentication based on Quality Measures. Pattern Recognition, Vol.38(5), pp.777-779, 2005.

49. L. Flom, A. Safir. Iris recognition system. US Patent 4641394, 1987.

50. N. Friedman, D. Geiger, M. Goldszmidt. Bayesian network classifiers, Proceedings of Machine Learning. Vol.29, pp.131–163,1997.

51. R. W. Frischholz, U. Deickmann. BioID: A multimodal biometric identification system," IEEE Computer, Vol. 33(2), pp.64-68, 2000.

52. J. Galbally, F. Alonso-Fernandez, J. Ortega-Garcia. A high performance fingerprint liveness detection method based on quality related features. Future Generation Computer Systems, Vol. 28(1), pp. 311-321, 2012.

53. J. Garcia-Nieto, E.G. Talbi, E. Alba, L. Jourdan. A comparison between Genetic Algorithm and PSO approaches for Gene selection and classification of Microarray data. ACM (GECCO-07), pp. 427–429. 2007,

54. G. Giacinto, F. Roli. Dynamic classifier selection based on multiple classifier behaviour. Pattern Recognition, Vol.34, pp.1879–1881, 2001.

55. C. Gruber, T. Gruber, S. Krinninger, B. Sick; "Online signature verification with support vector machines based on LCSS kernel functions. IEEE Transactions on Systems, Man, and Cybernetics, Part B, Vol. 40(4),pp. 1088-1100, 2010.

56. J.W. Grzymala-Busse. MLEM2 rule induction algorithms: With and without merging intervals. Studies in Computational Intelligence (SCI), Springer-Verlag, Vol. 118, pp. 153-164, 2008.

57. G.K. Gupta. Introduction to Data Mining with Case Studies. Prentice-Hall of India, 2006.

58. M. A. Hall, L. A. Smith. Practical feature subset selection for machine learning," Proceedings of the 21st Australian Computer Science Conference, Springer-Verlag, pp. 181-191, 1998.

59. M. Hall. Correlation-based Feature Selection for Machine Learning. PhD thesis, The University of Waikato, 1999.

60. M. Hall. Correlation-based feature selection for discrete and numeric class machine learning," Proceedings of 7th Intentional Conference on Machine Learning, Stanford University, CA. Morgan Kaufmann Publishers., 2000.

61. L. Hansen, P. Salamon. Neural networks ensembles. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.12(10), pp.993–1001, 1990.

62. A.E. Hassanien, H. Own. Rough sets for prostate patient analysis. Proceedings of International Conference on Modeling and Simulation (MS2006), 2006.

63. S. Haykin. Neural network – A comprehensive foundation. (2nd edition). Prentice Hall, 1999.

64. History of Fingerprinting. FINGERPRINTING. Home page, http://www.fingerprinting.com/history-of-fingerprinting.php, last visited 30th January 2011.

65. T.K. Ho, J.J. Hull, S.N. Srihari. Combination of decisions by multiple classifiers. In: H. Baird, H. Bunke, K. Yamamoto (Eds.), Structured Document Image Analysis, Springer-Verlag, pp.188–202, 1992.

66. T. Ho, J. Hull, S. Srihari. Decision combination in multiple classifier systems. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.16 (1), pp.66-75, 1994.

67. T.K. Ho. Data complexity analysis for classifier combination. Proceedings of the 2nd International Workshop of Multiple Classifier System, Lecture Notes in Computer Science, Springer-Verlag, Cambridge, UK, pp.53–67, 2001.

68. L. Hong, A. K. Jain, S. Pankanti. Can Multibiometrics Improve Performance?. Proceedings of IEEE Workshop on Automatic Identification Advanced Technologies (AutoID), New Jersey, USA, pp. 59-64, 1999.

69. C.W. Hsu, C. Chang, C. Lin. A Simple Decomposition Method for Support Vector Machines. Machine Learning, Vol.46, pp.291-314, 2002.

70. S. Huang. Survey of particle swarm optimization algorithm. Computer Engineering and Design. Vol.30(8), pp.1977-1980, 2009.

71. A. Jain , A. Ross. Fingerprint mosaicking. Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), Vol.4, pp.4064-4067, 2002.

72. A. K. Jain, A. Ross, S. Pankanti. Biometrics: A tool for information security. IEEE Transaction on Information Forensics and Security, Vol.1(2), pp.125-143, 2006.

73. A. Jain, U. Uludag. Hiding fingerprint minutiae in images. Proceedings of Third Workshop on Automatic Identification Advanced Technologies, pp. 97-102, 2002.

74. A. Jain, P. Flynn, A. Ross. Handbook of Biometrics. Springer, New York, 2008.

75. A. K. Jain, R. Duin, J. Mao. Statistical Pattern Recognition: A Review. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.22(1), pp. 4-37, 2000.

76. A. Jain, K. Nandakumar, A. Ross. Score normalization in multimodal biometric systems. Pattern Recognition, Vol.38, pp. 2270-2285, 2005.

77. X.Y. Jing, Y.F. Yao, J.Y. Yang, M. Li, D. Zhang. Face and palmprint pixel level fusion and kernel DCVRBF classifier for small sample biometric recognition. Pattern Recognition, Vol. 40(3), pp.3209–3224, 2007.

78. T. Joachims. A probabilistic analysis of the rocchio algorithm with tfidf for text categorization. Proceedings of the 14th International Conference on Machine Learning, pp. 143-151,1997.

79. M. Jordon, R. Jacobs. Hierarchical mixtures of expert and the EM algorithm. Neural Computing, pp.181–214, 1994.

80. A. K. Jain, A. Ross, and S. Prabhakar. An Introduction to Biometric Recognition. IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Image- and Video-Based Biometrics, Vol.14(1), pp.4–20, 2004.

81. A. Kale, A. Roy-chowdhury, R. Chellappa, "Fusion of gait and face for human identification. Proceedings of Acoustics, Speech, and Signal Processing, Montreal, (ICASSP), Vol.5, pp.901-904, 2004.

82. H. Kang, B. Lee, H. Kim, D. Shin, J. Kim. A study on performance evaluation of the liveness detection for various fingerprint sensor modules. Proceedings of KES, pp. 1245–1253, 2003.

83. J. Kennedy and R. Eberhart. Particle swarm optimization. Proceedings of the IEEE International Conference on Neural Networks , Vol.4, Perth, Australia, pp.1942–1948,1995.

84. J. Kennedy, R.C. Eberhart. Swarm intelligence. Morgan Kaufmann, 2001.

85. J. Kennedy, R. Eberhart. A discrete binary version of the particle swarm algorithm. Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Vol.5, pp. 4104–4108,1997.

86. A. Kholmatov, B. Yanikoglu. Identity authentication using improved online signature verification method. Pattern Recognition Letters, pp. 2400-2408, 2005.

87. N. Kingsbury. Image processing with complex wavelets. Phil. Trans. Royal Society London, Vol.357, pp. 2543–2560, 1999.

88. G.V. Kiran, R.S.R. Kunte, S. Samuel. On line signature verification system using probabilistic feature modelling. Proceedings of International Symposium on Signal Processing and Its Application (ISSPA), pp. 355–358, 2001.

89. J. Kitter , F. Alkoot. Sum versus vote fusion in multiple classifier systems. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.25, pp.110-115, 2003.

90. J. Kittler, M. Hatef, R.P.W. Duin, J. Matas. On Combining Classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.20(3), 1998.

91. J. Kittler, M. Messer. Fusion of multiple experts in multimodal biometric personal identity verification systems. Proceedings of the 12th IEEE Workshop on Neural Networks, Guildford (UK), pp. 3–12,2002.

92. J. Kludas, E. Bruno, S. Marchand-Maillet. Information fusion in multimedia information retrieval. Proceedings of the 5th international Workshop on Adaptive Multimedia Retrieval (AMR), Paris, France, 2007.

93. G. Koltzsch. Biometrics – Market Segments and Applications," Journal of Business Economics and Management, Vol.8(2), pp. 119-122, 2007.

94. J. Komorowski, L. Polkowski, A. Skowron. Rough sets: A tutorial. Lecture Notes for ESSLLI'99: the fifth European Summer School in Language, Logic and information, Utrecht, Holland, 1999.

95. W. Kong, D. Zhang. Accurate iris segmentation based on novel reflection and eyelash detection model. Proceedings of 2001 International Symposium on Intelligent Multimedia, Video and Speech Processing, Hong Kong, 2001.

96. S. Krawczyk. User authentication using on-line signature and speech. Master's thesis, Michigan State University, Department of Computer Science and Engineering, 2005.

97. A. Kumar, D. C. M. Wong, H. C.Shen1, A. K. Jain. Personal verification using palmprint and hand geometry biometric. Proceedings of the 4th International

Conference on Audio- Video-Based Biometric Person Authentication, J. Kittler and M. Nixon, Eds., 2003, Vol. LNCS 2668, pp. 668–678.

98. A. Kumar, D.C.M. Wong, H.C.Shen, A.K. Jain. Personal authentication using hand images. Pattern Recognition Letters, Vol.27(13), pp.1478–1486,2006.

99. L. Kuncheva. Switching between selection and fusion in combining classifiers: an experiment. IEEE Transactions on System and Man Cybernetic—Part B, Vol.32 (2), pp.146–156, 2002.

100. L. I. Kuncheva. That elusive diversity in classifier ensembles. Pattern Recognition and Image Analysis, Vol.2652, pp. 1126–1138, 2003.

101. P. Langley, S. Sage. Scaling to domains with irrelevant features. Computational Learning Theory and Neural Learning Systems, R. Greiner (Editor), Vol.4, MIT Press 1994.

102. P. Langley, S. Sage. Induction of selective Bayesian classifiers. Proceedings of the 10th International Conference of Uncertainty in Artificial Intelligence, Seattle, W. A., Morgan Kaufmann, 1994.

103. F. Leclerc, R. Plamondon. Signature verification: The state of the Art --1989-1993. International Journal of Pattern Recognition and Artificial Intelligence, Vol. 8(3), pp. 643- 660, 1994.

104. L.L. Lee. Neural approaches for human signature verification. Proceedings of Third International Conference on Signal Processing, Vol. 2, pp. 1055-1058, 1996.

105. L.L. Lee. Neural approaches for human signature verification. Proceedings of the Third International Conference on Signal Processing, pp. 1346–1349, 1996.

106. H. Lei, V. Govindaraju. A study on the consistency of features for on-line signature verification. Structural, Syntactic, and Statistical Pattern Recognition, Springer-Berlin/Heidelberg, pp.444-451, 2004.

107. D. Lejtman, S. George. On-line Handwritten Signature Verification Using Wavelets and Back-Propagation Neural Networks. Proceedings of the 6th International Conference on Document Analysis and Recognition, pp.596-598, 2001.

108. D.Z. Lejtman, S.E. George. On-line handwritten signature verification using wavelets and back-propagation neural networks. Proceedings of the 6th International Conference on Document Analysis and Recognition, Seattle, pp. 596-598, 2001.

109. LG. home page, http://www.lgiris.com/ps/technology/index.htm, last visited 30th September 2011.

110. J. Li, N. Cercone. A rough set based model to rank the importance of association rules. Proceedings of the 7th International Workshop on New Directions in Rough Sets, Data Mining, and Granular-Soft Computing (RSFDGrC ’99),

in:Lecture Notes in Computer Science, Springer, Berlin, Germany, pp. 109–118, 2005.

111. S. Lim, K. Lee, O. Byeon, T. Kim. Efficient iris recognition through improvement of feature vector and classifier. ETRI Journal, Vol. 23(2), pp.1–70, 2001.

112. X. Liu, T. Chen. Geometry-assisted statistical modeling for face mosaicing. Proceedings of IEEE ICIP, Vol.2, pp. 883–886, 2003.

113. J. Liu, Q. Hu, D. Yu. A comparative study on rough set based class imbalance learning. International Journal of Knowledge Based Systems, Vol.21, pp.753–763, 2008.

114. L. Ma, T. Tan, Y. Wang, D. Zhang. Efficient Iris Recognition by Characterizing Key Local Variations. IEEE Transactions on Image Processing, Vol.13(6), pp. 739-750, 2004.

115. L. Ma, Y. Wang, T. Tan. Iris recognition using circular symmetric filters" , Proceedings of the International Conference on Pattern Recognition, Vol.2, pp.414–417, 2002.

116. D. Maltoni, D. Maio, A. K. Jain, S. Prabhakar. Handbook of Fingerprint Recognition. Springer-Verlag, 2003.

117. L. Masek. Recognition of human iris patterns for biometric identification. Master's thesis, University of Western Australia, 2003.

118. MATLAB help, version R2007b.

119. T. Matsumoto, H. Matsumoto, K. Yamada, S. Hoshino. Impact of artificial 'gummy' fingers on fingerprint systems. Proceedings of SPIE, Vol.4677, pp. 275–289, 2002.

120. A. McCallum and K. Nigam. A comparison of event models for naive bayes text classification. Proceedings of the AAAI-98 workshop on Learning for Text Categorization,1998.

121. J. Meaney. History of Fingerprints Timeline, home page http://www.fingerprintamerica.com/fingerprinthistory.asp, last visited 18th May 2011.

122. D. Michie, D.J. Spiegelhalter, C.C. Taylor. Machine Learning, Neural and Statistical Classification. Ellis Horwood Ltd., London, UK, 1994.

123. T.M. Mitchell. Machine Learning. McGraw-Hill, New York, 1997.

124. D. Muramatsu, M. Kondo, M. Sasaki, S. Tachibana, T. Matsumoto. A Markov chain Monte Carlo algorithm for bayesian dynamic signature verification. IEEE Transactions on Information Forensics and Security, Vol.1(1), pp.22–34,2006.

125. E. N. Marieb. Human Anatomy and Physiology. Benjamin-Cummings Publishing Company, 7th edition, 2007.

126. V.S. Nalwa. Automatic on-line signature verification. Proceedings of IEEE, Vol.85 (2), pp.215–239,1997.

127. K. Nandakumar. Multibiometric Systems: Fusion Strategies and Template Security. PhD thesis, Department of Computer Science and Engineering, Michigan State University, 2008.

128. K. Nandakumar. Integration of Multiple Cues in Biometric Systems. Master thesis, Department of Computer Science and Engineering, Michigan State University, 2005.

129. L. Nanni, A. Lumini. A novel local on-line signature verification system. Pattern Recognition Letters, Vol.29 (5), pp. 559–568, 2008.

130. L. Nanni, E. Maiorana, A. Lumini, P. Campisi. Combining local, regional and global matchers for a template protected on-line signature verification system. Expert System Application, Vol.37 (5), pp.3676–3684, 2010.

131. S. P. Narote, A. S. Narote, L. M. Waghmare. An Automated Segmentation Method for Iris Recognition. Proceedings of the International IEEE Conference TENCON, 2006.

132. NIST Report to the United States Congress. Summary of NIST Standards for Biometric Accuracy, Tamper Resistance, and Interoperability. Available at ftp://sequoyah.nist.gov/pub/nist_internal_reports/NISTAPP_Nov02.pdf, September 2011.

133. R. Noori, M.A. Abdoli, A. Ameri, M. Jalili-Ghazizade. Prediction of municipal solid waste generation with combination of support vector machine and principal component analysis: a case study of Mashhad. Environmental Program Substantial Energy, Vol.28, pp. 249–258, 2009.

134. J. Ortega, D. Simon, M. Faundez, V. Espinosa, A. Satue, I. Hernaez, J.-J. Igarza, C. Vivaracho, Q.-I. Moro, "MCYT: A multimodal biometric database. Proceedings COST-275 Biometric Recognition Workshop, Rome, Italy, pp.123–126, 2002.

135. H.S. Own, W. Al-Mayyan, H. Zedan. Biometric-Based Authentication System Using Rough Set Theory. RSCTC 2010, pp. 560-569, 2010.

136. C. Oyster. "The Human Eye Structure and Function. Sinauer Associates, 1999.

137. L. Pan, M. Xie. Research on iris image preprocessing algorithm. IEEE International Symposium on Machine Learning and Cybernet, Vol.8, pp.5220–5224, 2005.

138. K.E. Parsopoulos, M.N. Vrahatis. Particle Swarm Optimization and Intelligence: Advances and Applications. USA, 2010.

139. Z. Pawlak, "Rough Sets. International Journal of Computer and Information Science, Vol.11, pp. 341-356, 1982.

140. Z. Pawlak, A. Skowron. Rudiments of rough sets. Information Sciences, Vol.177(1), pp. 3–27, 2007.

141. B. Pierscionek, S. Crawford, B. Scotney. Iris recognition and ocular biometrics- the salient features. Proceedings of Machine Vision and Image Processing Conference IMVIP '08. International, pp. 170-175, 2008.

142. C.E. Pippin. Dynamic signature verification using local and global features. Technical report, Georgia Inst. Inform. Technology, Atlanta, 2004.

143. R. Plamondon, G. Lorette. Automatic signature verification and writer identification- the state of the art. Pattern Recognition, Vol. 22, pp. 107–131,1989.

144. R. Plamondon, S.N. Srihari. On-line and off-line handwriting recognition: A Comprehensive Survey. IEEE Transactions on Pattern Matching and Machine Intelligence, Vol. 22(1), pp. 63-84, 2000.

145. N. Poh, S. Bengio. Database, protocols and tools for evaluating score-level fusion algorithms in biometric authentication. Pattern Recognition, Vol.39, pp.223–233, 2006.

146. N. Poh, J. J. Korczak. Hybrid Biometric Person Authentication Using Face and Voice Features. Proceedings of the Third International Conference on Audio- and Video-Based Biometric Person Authentication, pp. 348-353, 2001.

147. H. Proença, L.A. Alexandre. Iris Recognition: Measuring Feature's Quality for the Feature Selection in Unconstrained Image Capture Environments. IEEE Proceedings of the 2006 International Conference on Computational Intelligence for Homeland Security and Personal Safety-CIHSPS 2006, Alexandria, U.S.A ,Vol. 1, pp.35-40, 2006.

148. J. R. Quinlan. C4.5: Programs for machine learning. Morgan Kaufmann, Los Altos, California, 1993.

149. H. Raafat, A.S. Tolba, W. Almayaan. A glove-based system for automated signature identification using neural networks. The Fifth International Conference on Systems Analysis, and Synthesis, ISAS '99, 1999.

150. N. Ratha, J. Connell, R. Bolle. Image mosaicing for rolled fingerprint construction. 14th International Conf on Pattern Recognition, pp.1651-1653, 1998.

151. A. Rattani , M. Tistarelli. Robust multimodal and multiunit feature level fusion of face and iris biometrics. International Conference of Biometrics, Springer, pp.960–969,2009.

152. H. T. F. Rhodes. Alphonse Bertillon, Father of Scientific Detection. Abelard-Schuman, New York. 1956.

153. N. Ritter, J. Cooper. Locating the iris: A first step to registration and identification. Proceedings of the 9th IASTED International Conference on Signal and Image Processing, pp. 507-512, 2003.

154. A. Ross, K. Nandakumar, A. K. Jain. Handbook of Multibiometrics. Springer, 2006.

155. A. Ross, S. Dass, A. Jain. Fingerprint warping using ridge curve correspondences. IEEE Transactions on Pattern Analysis and Machine Intelligence Vol.28(1), pp. 19–30, 2006.

156. A. Ross, S. Shah, J. Shah. Image Versus Feature Mosaicing: A Case Study in Fingerprints. Proceedings of SPIE Conference on Biometric Technology for Human Identification, Vol. 6202, pp. 1–12, 2006.

157. A. Ross, A. Jain. Multimodal biometrics: an overview. Proceedings of the 12th European Signal Processing Conference (EUSIPCO), (Vienna, Austria), pp.1221-1224, 2004.

158. A. Ross, A. Jain. Information Fusion in Biometrics. Pattern Recognition Letters, Special Issue on Multimodal Biometrics, Vol. 24, pp. 2115-2125, 2003.

159. A. Ross, R. Govindarajan. Feature level fusion using hand and face biometrics. Proceedings of SPIE Conference on Biometric Technology for Human Identification, pp. 196–204, 2004.

160. C. S. Burrus, R. A. Gopinath, H. Guo. Introduction to Wavelets and Wavelet Transform. Prentice Hall, Englewood Cliffs, NJ, 1997.

161. C. Sanderson, K. K. Paliwal. Information Fusion and Person Verification Using Speech and Face Information. IDIAP-RR, pp.02-33, 2003.

162. M. Schumacher, N. Hollander, W. Sauerbrei. Resampling and cross-validation techniques: a tool to reduce bias caused by model building?. Statistics in Medicine, Vol.16, pp.2813–2827,1997.

163. I.W. Selesnick, R.G. Baraniuk, N.C Kingsbury. The dual-tree complex wavelet transform. IEEE Signal Processing Magazine, Vol. 22(6), pp.123–151, 2005.

164. G. Shafer. A Mathematical Theory of Evidence. Princeton University Press, Princeton, pp.297,1976.

165. M.M. Shafiei, H.R. Rabiee. A new online signature verification algorithm using variable length segmentation and hidden Markov models. Proceedings of the Seventh International Conference on Document Analysis and Recognition, Vol.1, pp.443–446, 2003.

166. G. Shakhnarovich, T. Darrell, P. Indyk. Nearest-Neighbor Methods in Learning and Vision: Theory and Practice. MIT Press, 2006.

167. L. Shapiro, G. Stockman. Computer Vision. Prentice-Hall, Inc. 2001.

168. Y. Shi, R. C. Eberhart. A modified particle swarm optimizer. Proceedings of the IEEE International Conference on Evolutionary Computation, pp.69-73,1998.

169. D. Slezak. Various approaches to reasoning with frequency-based decision reducts: a survey. In: L. Polkowski, S. Tsumoto, T.Y. Lin (Eds.), Rough Sets in Soft Computing and Knowledge Discovery: New Developments, Physical Verlag, 2000.

170. R. Snelick, M. Indovina, J. Yen, A. Mink. Multimodal biometrics: issues in design and testing. Proceedings of Fifth International Conference on Multimodal Interfaces, Vancouver, Canada, pp. 68–72, 2003.

171. R. Snelick, U. Uludag, A. Mink, M. Indovina, A. Jain. Large Scale Evaluation of Multimodal Biometric Authentication Using State-of-the-Art Systems," IEEE Trans. Pattern Analysis and Machine Intelligence, Vol.27(3), pp. 450-455, 2005.

172. M.P. Song, G.C. Gu. Research on particle swarm optimization: a review. Proceedings of 2004 International Conference on Machine Learning and Cybernetics, Vol.4, pp. 2236–2241, 2004.

173. M. Song, G. Gu. Research on particle swarm optimization: A review. Proceedings of the third international conference on machine learning and cybernetics, Shanghai, 2004.

174. Stan Li (Editor). Encyclopedia of Biometrics. Elsevier Publisher, 2009.

175. P. Stavroulakis, M. Stamp. Handbook of Information and Communication Security. Springer, 2010.

176. Z. Sun, T. Tan. Ordinal Measures for Iris Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.31(12), pp.2211-2226, 2009.

177. Topaz Systems Inc., home page, http://www.topazsystems.com, last visited 28th September 2011.

178. J. Thornton, M. Savvides, M. Kumar, B.V.K. Vijaya. An evaluation of iris pattern representations. Biometrics: Theory, Applications, and Systems, 2007.

179. C. Tisse , L. Martin , L. Torres ,M. Robert. Person Identification Technique Using Human Iris Recognition. Proceedings Vision Interface, pp. 294-299, 2002.

180. K.A. Toh, X.D. Jiang, W.Y. Yau. Exploiting global and local decisions for multi-modal biometrics verification. IEEE Transactions on Signal Processing, Vol.52(10), pp. 3059–3072, 2004.

181. V.N. Vapnik. The Nature of Statistical Learning Theory. Springer-Verlag, New York, 1999.

182. T. Wang, T. Tan, A. Jain. Combining face and iris biometrics for identity verification. Proceedings of the 4th International Conference on Audio- Video-Based Biometric Person Authentication, J. Kittler and M. Nixon, Eds., 2003, Vol. LNCS 2688, pp. 805–813.

183. X.Y. Wang, J. Yang, X. Teng, W. Xia, R. Jensen. Feature selection based on rough sets and particle swarm optimization. Pattern Recognition Letters, Vol.28(4), pp.459–471,2007.

184. J. Wayman, A. Jain, D. Maltoni, D. Maio (Editors.). Biometric Systems: Technology, Design and Performance Evaluation. Springer, 2005.

185. R.P. Wildes. Iris recognition: an emerging biometric technology. Proceedings of the IEEE, Vol.85(9), pp. 1348–63, 1997.

186. P. Xiuqin, X. Xiaona, L. Yong, C. Youngcun. Feature fusion of multimodal recognition based on ear and profile face. Proceedings SPIE-2008, 2008.

187. F. Yang, M. Paindavoine, H. Abdi, D. Arnoult. Fast Image Mosaicing for Panoramic Face Recognition. Journal of Multimedia, Vol. 1(2), pp. 14-20, 2006.

188. L. Yang, B. Widjaja, R. Prasad, Application of hidden Markov models for signature verification, Pattern Recognition, Vol.28 (2), pp.161–170,1995.

189. Y.Yao, X. Jing, and H. Wong, Face and palmprint feature level fusion for single sample biometric recognition. Nerocomputing, Vol.70(7-9), pp. 1582–1586, 2007.

190. D.Y. Yeung, H. Chang, Y. Xiong, S. George, R. Kashi, T. Matsumoto, G. Rigoll. SVC2004: First international signature verification competition. Proceedings of Third Workshop on Automatic Identification Advanced Technologies, pp.97–102, 2002.

191. M.F. Zanuy. On-line signature recognition based on VQ-DTW. Pattern Recognition, Vol.40 (3), pp. 981–992, 2007.

192. D. Zhang, X. Jing and J. Yang. Biometric Image Discrimination (BID) Technologies. IGP/IRM Press edition, 2006.

193. N. Zhong, A. Skowron. Rough sets based knowledge discovery process. International Journal of Applied Mathematics and Computer Science, Vol.11(3), Technical University Press, Poland, pp.101-117, 2001.